Noise detection thresholds after exposure to pulse-train sounds

Jarkko Lamminsalo

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 31.8.2016

Thesis supervisor:

Prof. Ville Pulkki

Thesis advisors:

Ph.D Catarina Hiipakka

D.Sc. (Tech.), Docent Okko Räsänen

Aalto University
School of Electrical Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Jarkko Lamminsalo

Title: Noise detection thresholds after exposure to pulse-train sounds

Date: 31.8.2016 Language: English Number of pages: 9+62

Department of Signal Processing and Acoustics

Professorship: Acoustics and Audio Signal Processing Code: S-89

Supervisor: Prof. Ville Pulkki

Advisors: Ph.D Catarina Hiipakka, D.Sc. (Tech.), Docent Okko Räsänen

An auditory aftereffect is an illusory perceptual experience occurring in the hearing system immediately after an exposure to a sound. There is very little research done on aftereffects, thus, many aspects regarding them are still unknown. One particular kind of auditory aftereffect, elicited by a pulse-train sound, has been described to temporarily change the timbre of consecutive sounds into "metallic". However, it is not known whether this altered timbre is also associated with other kinds of changes, such as changes in hearing thresholds.

A listening experiment was conducted in this thesis, with the goal of finding out whether the pulse-train aftereffect, also known as the Rosenblith aftereffect, alters the minimal perceivable sound pressure level, i.e. the detection threshold, of sounds in comparison to aftereffect-free cases. The detection thresholds of noise, harmonic noise and two different aftereffect-eliciting pulse trains were measured in an anechoic chamber using headphones and an adaptive tracking procedure. The test sound was preceded by a long exposure to one of the four sounds. The results show that there was no clear connection between the aftereffect and alterations in noise detection thresholds. However, indications of an aftereffect-related change in the pure tone detection threshold were found.

The obtained data also reveal phenomena outside the main goal. A sound that the subject was just exposed to, was generally detected at a lower threshold, in some cases up to 1 dB lower, than some other previously unknown sound. This is contradictory to the elevated threshold hypothesized by adaptation, and highlights the involvement of higher level hearing processes in the detection. The third finding concerns the recovery from the exposure. In some cases an unexpected bounce-like elevation in the detection threshold was found at around 2 minutes after the exposure.

Keywords: psychoacoustics, auditory aftereffect, pulse-train, noise

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Jarkko Lamminsalo

Työn nimi: Kohinan havaitsemiskynnykset pulssijonoäänille altistumisen jälkeen

Päivämäärä: 31.8.2016 Kieli: Englanti Sivumäärä: 9+62

Signaalinkäsittelyn ja akustiikan laitos

Professuuri: Akustiikka ja äänenkäsittely Koodi: S-89

Työn valvoja: Prof. Ville Pulkki

Työn ohjaajat: Ph.D Catarina Hiipakka, TkT, Dos. Okko Räsänen

Auditorinen jälkiefekti on illuusiomainen havaintokokemus, joka ilmenee kuulojärjestelmässä välittömästi tietynlaisille äänille altistumisen jälkeen. Jälkiefekteistä on tehty hyvin vähän tutkimusta, joten ne ovat vielä monelta osin tuntemattomia. Eräs tietty jälkiefekti, jonka pulssijonoääni synnyttää, on aiemmin kuvailtu väliaikaisena "metallisena" muutoksena altistuksen jälkeen kuultujen äänien sävyssä. On kuitenkin epäselvää, liittyykö muutos sävyssä myös muunlaisiin muutoksiin, esimerkiksi muutokseen kuulokynnyksessä.
Tässä diplomityössä suoritetun kuuntelukokeen päätavoitteena oli selvittää, muuttaako pulssijono -jälkiefekti, joka myös Rosenblith -jälkiefektinä tunnetaan, pienintä mahdollista äänenpainetasoa jolla ääniä havaitaan verrattuna tapauksiin ilman jälkiefektiä. Kohinan, harmonisen kohinan sekä kahden jälkiefektin synnyttävän pulssijonon havaintokynnykset mitattiin kaiuttomassa huoneessa käyttäen kuulokkeita ja adaptiivista menetelmää. Testattavaa ääntä edelsi pitkä altistus yhdelle näistä neljästä äänestä. Tulokset osoittavat etteivät jälkiefekti ja muutokset kohinan havaintokynnyksessä liity suoraan toisiinsa. Viitteitä jälkiefektiin liittyvistä muutoksista äänesten havaintokynnyksessä kuitenkin löydettiin.
Kerätty aineisto johti myös muihin löydöksiin. Ääni, jolle koehenkilö juuri altistettiin, havaittiin yleensä jopa 1 dB matalammalla kynnyksellä kuin ääni, jota koehenkilö ei ollut juuri kuullut. Tämä löydös on ristiriidassa adaptaatioteorian kanssa, jonka mukaan havaintokynnys kohoaisi altistuksen myötä. Löydös on osoitus siitä, että havaintokynnykseen vaikuttavat myös korkeamman tason kuuloprosessit. Kolmas diplomityön löydös liittyy palautumiseen altistuksesta. Joissakin tapauksissa nähtiin odottamaton yhtäkkinen nousu kuulokynnyksessä noin 2 minuuttia altistuksen jälkeen.

Avainsanat: psykoakustiikka, auditorinen jälkiefekti, pulssijono, kohina

Acknowledgements

This Master's thesis has been done in the Department of Signal Processing and Acoustics at Aalto University. The project was funded by the Academy of Finland project titled Computational Modeling of Language Acquisition (No. 269279).

I want to thank my instructors Catarina Hiipakka and Okko Räsänen for their invaluable guidance and help. Another big thank you goes to a man full of ideas, my supervising professor Ville Pulkki. Apart from his importance in keeping the atmosphere cheerful, he offered me the chance to share my knowledge on acoustics as a teacher at the University of the Arts. All my co-workers also deserve my deepest gratitude, especially Henri Pöntynen who shared his practical experiences on writing the thesis and Alessandro Altoè who provided some papers and graphs.

In the beginning of this project I bought myself a 36-string home kantele which deserves special thanks. Learning to play it has counterbalanced the whole Master's thesis process in a creative way. The final thank you goes to my family and friends for supporting me in my studies.

Otaniemi, 31.8.2016

Jarkko Lamminsalo

Contents

Abbreviations

AFC	Alternative forced choice
ANOVA	Analysis of variance
dBHL	dB Hearing Level
DT	Detection threshold
PTA	Pure tone average
RMS	Root mean square
SPL	Sound pressure level
TTS	Temporary threshold shift

Stimulus abbreviations

H	harmonic noise
N	noise
R	Rosenblith pulse train
Rf	Frequency modulated Rosenblith pulse train

List of Figures

List of Tables

1 Introduction

This thesis aims at widening the knowledge on auditory aftereffects. In the psychophysical literature the term aftereffect refers to an illusory perceptual experience that occurs immediately after the exposure to a physical stimulus, e.g. image or sound. One famous example of an aftereffect in the visual domain is the waterfall effect, in which fixed objects, such as rocks, look like they are moving after watching a waterfall for a while [1]. Probably the most well-known auditory aftereffect is the Zwicker tone, which is an illusion of a pure tone heard after listening to a specific kind of noise [2]. Although aftereffects exist both in sight and hearing, there are less examples of auditory aftereffects than there are of visual aftereffects. One explanation for this is that the auditory aftereffects require more specific and somewhat unnatural conditions to occur.

The literature on auditory aftereffects is sparse and many aspects, e.g. the mechanisms causing it, are still mostly unknown. The experiment conducted in this thesis aims at finding out whether a specific perceptual aftereffect, the pulse-train effect, alters the detection of sounds presented at around the minimum detectable sound pressure level (SPL). The pulse-train effect is called the Rosenblith aftereffect throughout the thesis, referring to the first paper published on it ([3]). The stimuli used in the experiment included several types of wide-band sounds: two pulse trains known to elicit the effect, and noise and harmonic noise known not to cause the effect. Psychophysical methods were used to measure the noise detection thresholds of the stimuli after one minute exposure to them. In this thesis the term (noise) detection threshold (DT) without further specifications refers to the detection threshold of wide-band sounds. This should not be confused with e.g. the DT of a pure tone and the DT of modulation.

There were three goals in the study:

1. To find out whether the DT is different after exposure to an aftereffect-eliciting sound compared to an aftereffect-free sound. In particular, is there a temporary threshold shift associated with the aftereffect?
2. To find out whether listening to a specific sound alters the DT of that sound only. For example, is it easier or harder to detect a pulse train after listening to it for a long time?
3. To analyze the time course of the DTs after the exposure to the stimuli.

This thesis consists of six chapters. Chapter 2 dives into the human hearing system and the psychophysical methods for studying it. Known auditory aftereffects are also introduced. Chapter 3 gives higher level motivation for the experiment. The conducted experiment is explained in detail in chapter 4. Chapter 5 draws conclusions about the obtained results suggesting improvements and future research topics. Chapter 6 summarizes the findings of this study.

2 Background

This section introduces the background necessary to understand the auditory aftereffects and how they can be studied. Section 2.1 introduces the hearing system and some of its properties in order to help understand the potential mechanisms behind the effects. The hearing research methods used to study the effects are discussed in section 2.2. The most important perceptual aspects related to the effects are introduced in section 2.3. The most well-known and studied auditory aftereffects are presented in 2.4.

2.1 The human auditory system

The human hearing system is capable of analyzing and extracting information, such as pitch, timbre and sound source direction, from a sound arriving to the two ears [4, Chapter 5]. The system is shown in Fig. 1. It can be divided into the peripheral and the central hearing. The peripheral hearing consists of the ears, which convert the longitudinal vibration of air, i.e. sound, to neural signals. The signals are processed further in the central hearing system and interpreted in the auditory cortex, located in the temporal lobe of the brain. The whole chain from ear to a subjective observation in the consciousness is called the hearing.

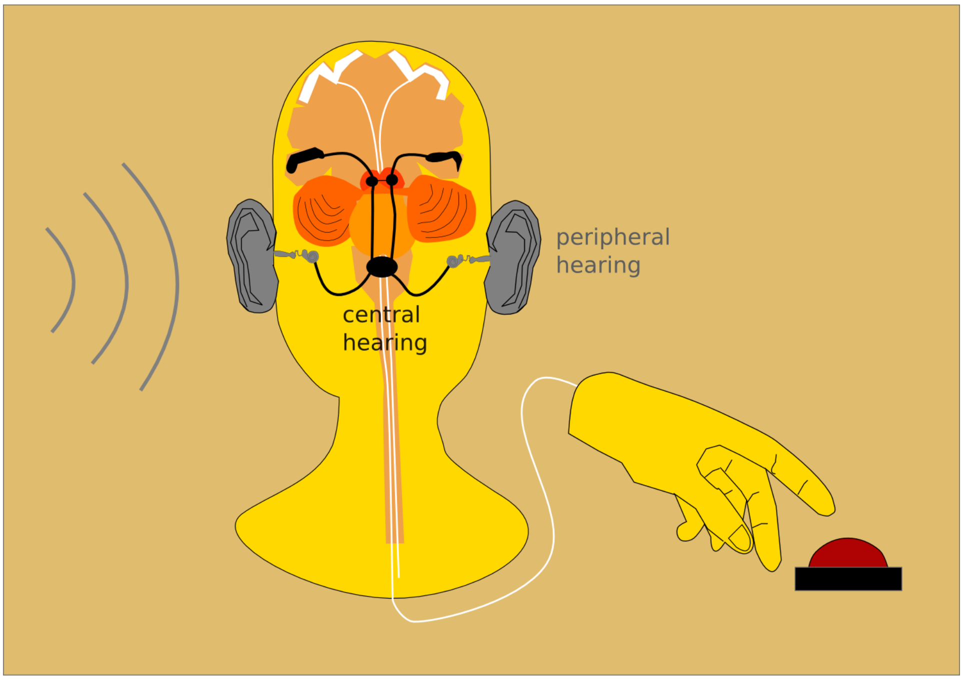

Figure 1: Schematic picture of how the perception of a sound leads to an action. Vibration of the air is received by the peripheral hearing system, marked in grey. The sound is transformed into a neural form and processed further in the central hearing system, marked in black. The sound is interpreted in the brain. The motor system, marked in white, generates the response needed e.g. for a button press.

The average human hearing range is shown in Fig. 2. The figure shows both the frequency and the dynamic ranges, i.e. how low or high sounds humans can perceive and at which levels the sounds can be perceived. The ears can process frequencies from 20 to 20000 Hz which corresponds to the wave length range from 20 m to 2 cm. The intensity ratio of the weakest and the loudest sound is more than 10^{12}. Both ranges are considered notably wide.

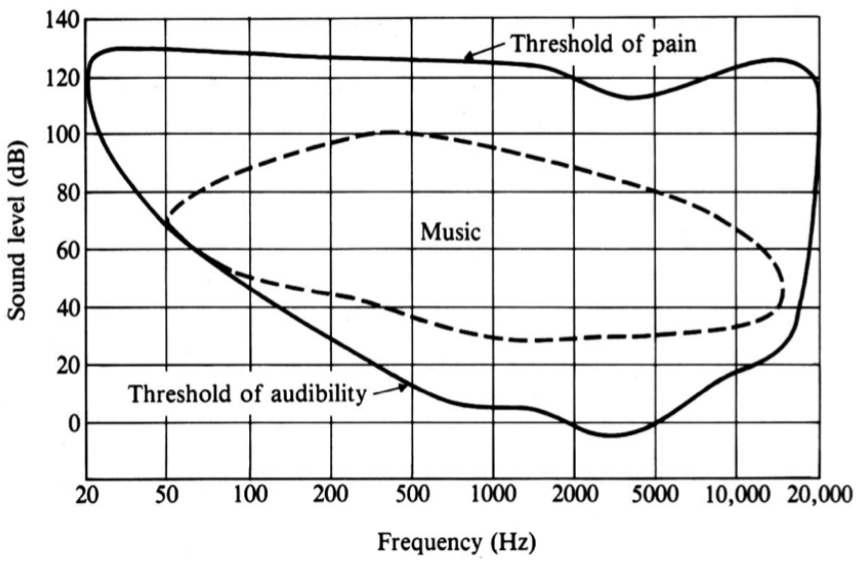

Figure 2: The human hearing range with the frequency and dynamic range of music drawn. The whole hearing range is rarely utilized in everyday life. Adopted from [4, p. 100].

2.1.1 Peripheral hearing

Peripheral hearing encompasses the ears which can be divided into three parts, as seen in Fig. 3: the outer, the middle and the inner ear [4, Chapter 5]. The outer ear consists of the pinna, the shape of which can differ rather much from individual to individual [5]. The pinna both collects and colours sound. The coloration is dependent of the direction of the sound source which is utilized in the brain as a cue for defining the direction of the sound source. The ear canal also colours the sound. It is a half-open pipe which is on average 2.5 cm long. This means that frequencies in the range of 2–4 kHz resonate and thus get acoustically amplified in the ear canal. Sound traveling through the ear canal vibrates the eardrum which is a border between the outer and the middle ear.

The vibrating eardrum makes the ossicles vibrate. They transfer the sound to the inner ear, amplifying the sound. The amplified sound enters the spiral shaped cochlea, shown in Fig. 4, through the oval window. There is a thin strip-like basilar membrane inside the cochlea, the end of which is called the helicotrema. The sound that enters the cochlea makes the basilar membrane vibrate according to its frequency

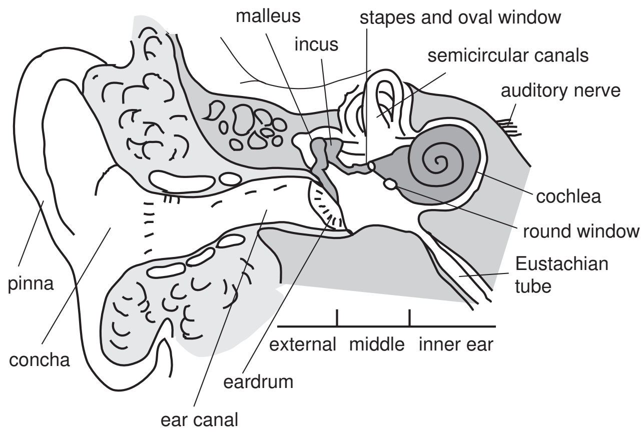

Figure 3: Structure of the ear. The sound travels from the outer ear through the middle ear into the inner ear. Adopted from [6].

content [4, Chapter 5]. Due to the varying mechanical properties, i.e. the width and the stiffness, of the membrane along its length, the lowest frequencies make the membrane vibrate the best closest to the helicotrema, whereas the highest frequencies cause vibrations near the oval window. However, also the low frequencies make the beginning of the membrane vibrate temporarily as the wave enters the cochlea and travels towards the helicotrema.

When a pure tone is played it causes a maximum displacement at one part of the membrane, each frequency having its maximum displacement in a different part along the length of the membrane [4, Chapter 5]. In this way each position in the membrane is associated with a frequency called the characteristic frequency. The displacement for four different characteristic frequencies is shown in Fig. 5. The displacement is asymmetric: the excitation at one frequency spreads more to the frequency regions above the characteristic frequency than below it. Also, the excitation spreads wider at lower frequencies (e.g. 50 Hz) than at higher frequencies (e.g. 1600 Hz).

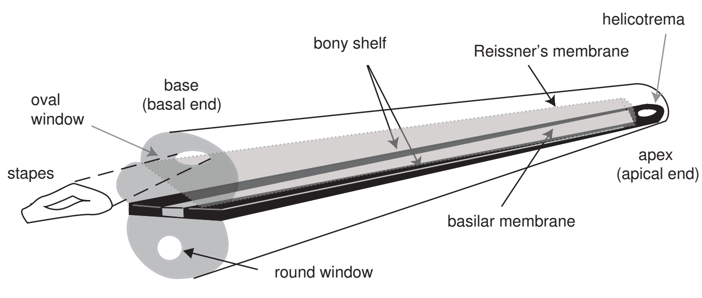

Figure 4: Structure of the cochlea. The cochlea has been straightened up for illustrational purposes. Adopted from [6].

There are about 20000 - 30000 hair cells along the basilar membrane [4, Chapter 5]. When the membrane is displaced the hair cells in the displaced part of the membrane are activated. In this way the cochlea performs spectral analysis, illustrated in Fig. 6. The neural signals from the different parts of the basilar membrane are carried in

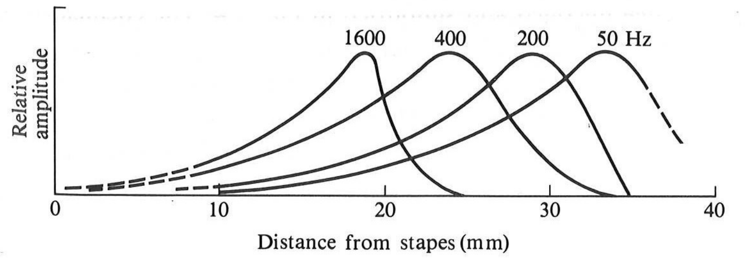

Figure 5: Displacement of the basilar membrane for various frequencies. Adopted from [4, p. 69].

separate nerve fibers which intertwine. Together they form the auditory nerve in which the signal from the cochlea is carried to the central hearing system as a series of pulses, e.g. neural spikes. The spikes correspond to the periods of a pure tone and correlate with the mechanical vibration of the basilar membrane up to 4–5 kHz. Different frequency and sound pressure ranges are carried in separate nerve fibres.

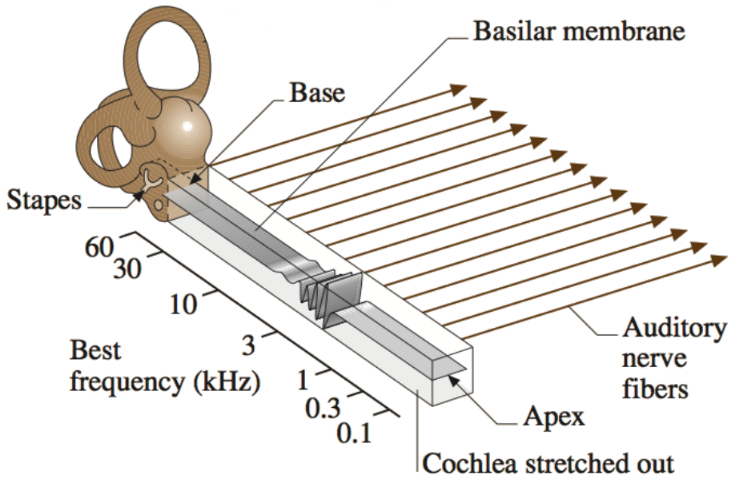

Figure 6: The frequency selectivity of the basilar membrane. Low frequencies make the membrane vibrate close to the end of the basilar membrane (apex), high frequencies close to the base. Sounds with a wide frequency spectrum make the whole membrane vibrate stimulating a greater amount of hair cells than a pure tone. Adopted from [7, p. 437].

The activations of the hair cells close to each other are processed in groups. These groups are called the critical bands and there are about 25 of them. A sound activating more than one critical band is perceived to be louder than a sound activating only one critical band even though both sounds have the same amount of energy. The ear processes the sound using critical bands very much in the same way as the eye processes light by dividing it into red, green and blue components. The critical bands are not equally spaced in frequency: there are more critical bands at lower frequencies.

Critical bands are some hundreds of Hzs wide in low frequencies whereas the width is some kHzs in higher frequencies. The frequency scale utilizing the critical bands is called the Bark scale. [4, Chapter 5]

2.1.2 Central hearing

The auditory pathway from the cochlea to the auditory cortex is a complex structure of ascending and descending paths [6]. A simplified version of the ascending pathways is shown in Fig. 7. The first processing station on the way to the cortex is the cochlear nucleus (CN) consisting of the ventral cochlear nucleus (VCN) and dorsal cochlear nucleus (DCN). The CN performs preamplification and preprocessing. The second station is the superior olivary complex (SOC) consisting of the medial superior olive (MSO) and the lateral superior olive (LSO). The SOC processes interaural differences enabling directional hearing. The third station is the inferior colliculus (IC) which receives signal both from the SOC and directly from the CN via the lateral lemniscus (LL). The task of the IC is e.g. to process spectral information and to integrate visual and head movement information to the auditory information. From this on the signal travels via the medial geniculate nucleus (MGN), located in the thalamus, to the primary auditory cortex where the processing is tonotopically organized. The higher the neurons in the auditory system the slower they typically react to changes in sound.

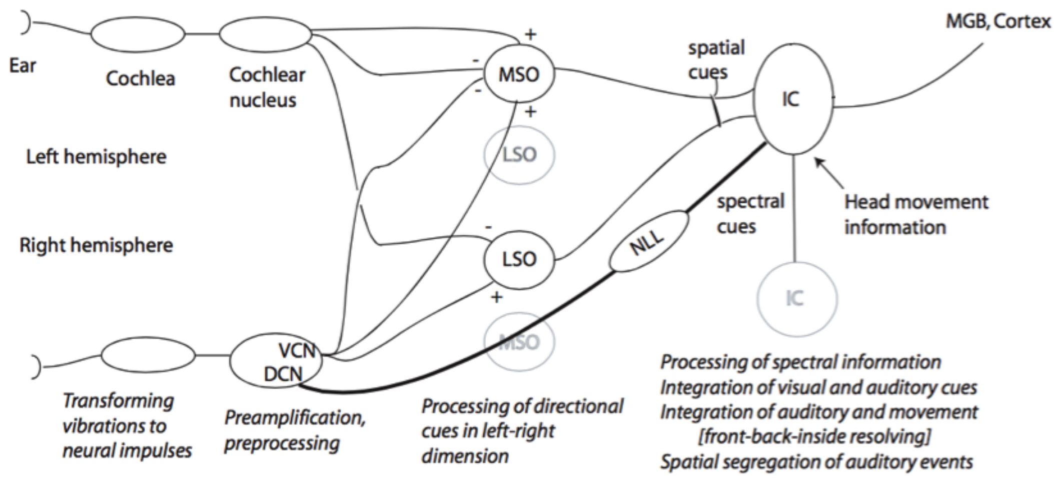

Figure 7: Ascending paths of the central hearing. Adopted from [6].

2.2 Hearing research methods

The hearing system can be examined with the help of psychophysics which is a systemic approach for studying the human sensory information processing, such as hearing, smell, vision or touch. Studying the auditory system with psychophysical methods is called the psychoacoustics. Compared to other study methods for studying the auditory system, such as direct neural measurements, psychoacoustic experiments are easier to conduct, non-invasive and pose no risk to the subjects [6, p. 133]. However, the downside is that the obtained data is often hard to analyse.

The way how psychoacoustics approaches the auditory system is shown in Fig. 8. In a psychoacoustic experiment the subject is played back a sound event s_i, for which the subject's conscious response is measured. The hearing system processes the sound event translating it into an auditory event h_i which is subjective. The translation follows the psychophysical function h. In an experiment the subject is asked to describe the auditory event, e.g. with numbers or words. This is typically done in a simplified manner by pressing buttons or adjusting sliders. The auditory event is translated into a description b_i which is the data obtained in the measurement. This translation is denoted as the description function b. Noise is added at all stages.

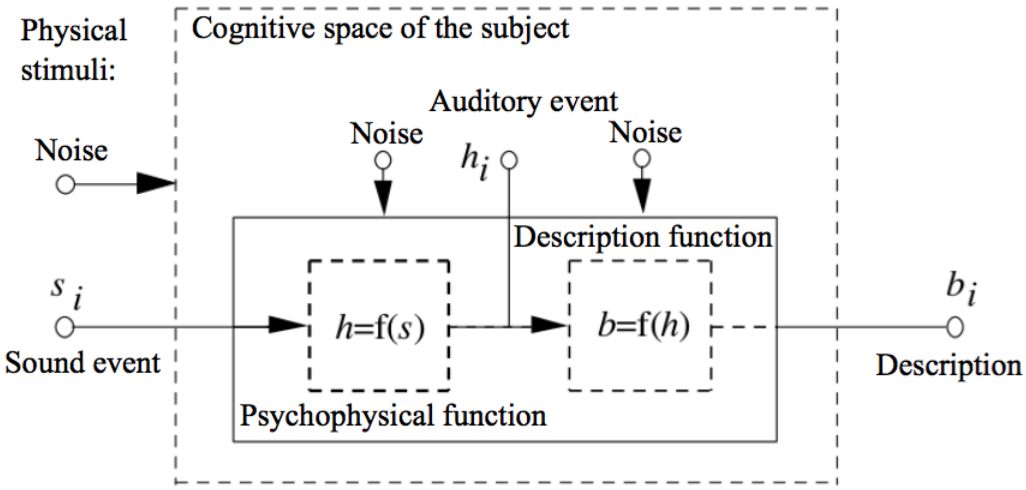

Figure 8: A block diagram showing how psychoacoustics models the measurement. Adapted from [8].

The auditory system is rather complex to model because it is non-linear and time-varying. In addition to the stimulus designed by the experimenter there are also many unintended factors, such as fatigue and concentration, that influence the system during a measurement [6, pp. 133–135]. A successful experimental design minimizes these noise factors. The results of a psychophysical measurement depend on the subjects, i.e. their past experiences, skills and prior knowledge [9, pp. 118–135]. Ideally the subjects would be naïve regarding the purpose of the experiment. The experiment design, for example the way how the subject is asked to respond, also affects the results.

The most important psychoacoustic research methods related to the study of this thesis are presented in the following sections. Prior to our actual experiment a pure-tone audiometry was conducted to all subjects. This is described in section 2.2.1. A simplest psychoacoustic experimental design for measuring responses to stimuli of different levels is the method of constant stimuli, introduced in section 2.2.2. The measurement of the DT in the actual experiment was based on a more advanced method, the transformed up-down method, proposed in [10]. The response logic was based on the Alternative Forced Choice (AFC) method. These can be found in sections 2.2.3 and 2.2.4. The Analysis of Variance (ANOVA), introduced in section 2.2.5, is an important statistical analysis method used widely in psychoacoustics.

2.2.1 Pure-Tone Audiometry

Pure-Tone Audiometry (ISO 8253-1:2010, [11]) measures absolute hearing thresholds for pure tones. Typically at least frequencies 250, 500, 1000, 2000 and 4000 Hz for both ears are measured. In the measurement a short pure tone is played through calibrated headphones in a quiet listening booth. The subject is given a button and instructed to press it when he or she hears the tone. The sound pressure level (SPL) of the pure tone is adjusted in an adaptive manner to find the threshold. The results are presented in an audiogram, such as the one in Fig. 9. Low background noise level is of great importance during the measurement since even white noise having a power spectral density of only -10 dB can mask sine tones [12].

The audiogram tells how much the absolute threshold at the tested frequencies and ears differ from the normal hearing threshold, referred to as the threshold of audibility previously in Fig. 2. The audiogram measures the absolute thresholds in the dB(HL) scale, in which the value 0 means that the threshold at the tested frequency is exactly at the normal hearing threshold. In practice the levels measured with audiometry fluctuate below and above the zero level, and this is considered normal. The levels greater than 20 dB indicate a hearing loss at the tested frequency for the tested ear.

2.2.2 Method of constant stimuli

The method of constant stimuli is a simple experimental design for measuring responses to stimuli [13, pp. 1215–1217]. In the method several stimulus levels, usually from five to nine, are chosen around the threshold value beforehand. All the stimuli are presented multiple times, usually more than 20 times, to all the subjects in random orders. The advantages are that the number of stimuli to be presented is known in advance and the obtained data covers a wide range of stimulus values, e.g. SPL levels. As a downside the appropriate stimulus range needs to be defined beforehand, e.g. by running preliminary tests. Also, because the set of stimulus levels is fixed, many trials are required to collect a sufficient amount of repetitions in the region of stimulus values where the effect takes place. This prolongs the experiment which might make it more difficult for the subjects to stay attentive.

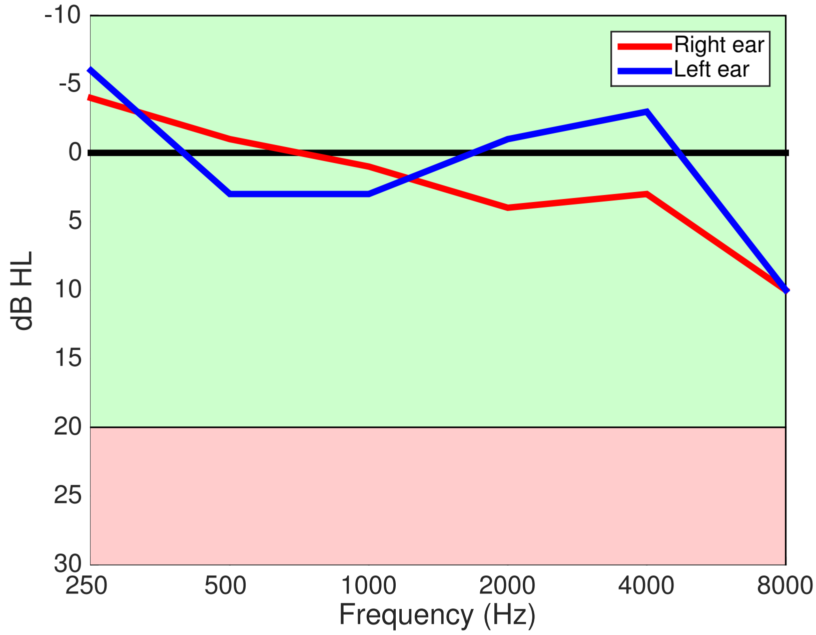

Figure 9: Results of a pure-tone audiometry plotted as an audiogram. dBHL level greater than 20 (the red region) indicates a hearing loss at that frequency for that ear.

2.2.3 Adaptive methods

The test duration can be reduced from the method of constant stimuli by adapting the presentation of the stimuli based on the responses of the subject. These kind of methods are called adaptive methods. The basic principle is to decrease the stimulus level after a positive response and increased it after a negative response. Thus, only the essential values around the threshold are tested. The test typically goes on until six or eight reversals, or turning points, have taken place [10]. The final threshold obtained from the test is the mean of the points where the reversals occurred.

The adaptive method can be modified further according to the needs of the experiment. One type of a transformed method is the so called "one-up two-down" method [10], in which one negative response is needed for the stimulus level to increase, go "up", and two consecutive positive responses for the stimulus level to decrease, go "down". A typical adaptive track from the experiment of this thesis with six turning points is shown in Fig. 10. The "one-up two-down" method can be preferred when the odds of getting a positive response are lower than the odds of getting a negative response. Its advantage over the most basic "one-up one-down" is the improved robustness against guessing. However, as a downside, more responses are needed for the adaptive track to converge, which increases the test duration.

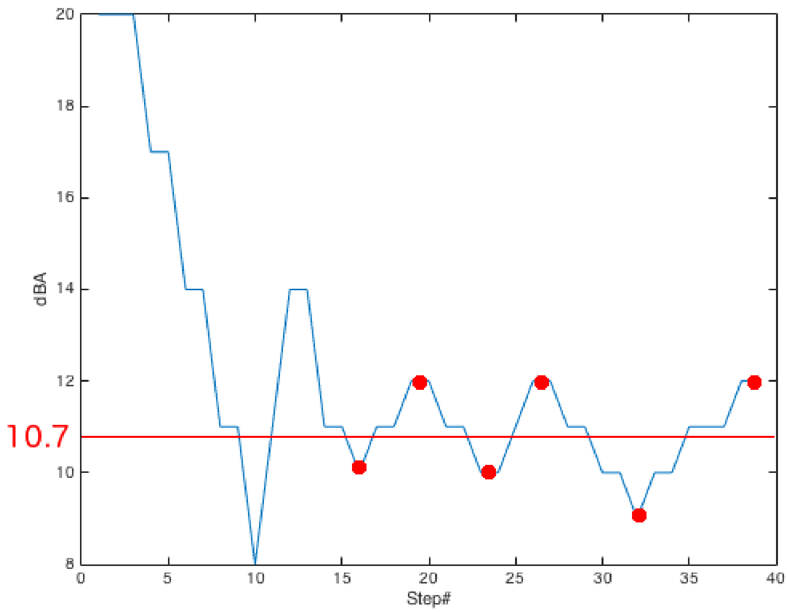

Figure 10: A typical adaptive track from the experiment of this thesis. The last six turning points are marked in red. The red line presents the final threshold which is the mean value of the last six turning points.

2.2.4 Alternative forced choice

Both in the method of constant stimuli and the adaptive the responses are collected after the presentation of a stimulus. One way of doing this is to ask the observer to respond, e.g. to press a button, if he or she perceived the stimulus. However, in that case the experimenter has no knowledge on the observer's internal criterion, i.e. the rule that he obeys in converting sensory information into overt responses [13, p. 1219].

The problem of the internal criterion can be avoided by using the Alternative Forced Choice (AFC) method for collecting the responses. In AFC the observer is given predefined choices and forced to respond on every trial. In fact, analysis of the AFC responses has revealed that observers can discern sounds so weak that they claim not to have detected. The amount of given choices can be customized. For example, if three choices are given, the method is called "3-AFC". The more choices there are, the more unlikely it becomes to give false positive responses – in other words, the method becomes more robust.

2.2.5 Analysis of variance

In the present thesis, the statistical analysis was carried out using the analysis of variance (ANOVA). It is a method for comparing the deviation between sample means of conditions to the random deviation within the samples and testing whether the magnitude of the ratio is higher than a certain critical value [9]. The results of the ANOVA can be used to infer whether the differences between two experiment conditions are caused by the differences (or no difference) intended by the experimenter or rather by random factors affecting both experiment conditions.

The data obtained in a measurement has to be inspected before analysis. The first important step is the detection and removal of outliers from the data [9, pp. 141–224]. If this is not done the ANOVA might return incorrect results. In many cases the outliers can be detected by visually inspecting the data. A general rule of thumb is to consider an observation an outlier if its value deviates more than 2.5 times the standard deviation. For more about the factors causing outliers, see [14].

The ideal data for ANOVA fulfills the following assumptions [15, p. 324]:

1. The data is drawn from a normally distributed population.
2. The variances in all experimental conditions are fairly similar, i.e. homogeneous.
3. Observations are independent.
4. The dependent variable is measured on at least an interval scale.

There are several ways of testing the assumptions [15, pp. 93–100]. In many cases a visual inspection of the plots can be sufficient. There are also tests with which to check some of the assumptions. Assumption 1, normality, can be tested e.g. with Shapiro-Wilk and Kolmogorov-Smirnov tests. Assumption 2, the homogeneity of variance, can be tested with the Levene's test, which compares the variances of two or more groups. Assumptions 3 and 4 are more difficult to test, thus, common sense is the best tool in evaluating whether they are violated.

Even if the data would not meet all the requirements it can still be analysed with the ANOVA. For instance, ANOVA is known to be quite robust against the violation of assumption 1 [16]. If assumption 2 is violated a corrected version of ANOVA, such as the Welch ANOVA, can be used to get valid results. The most serious problem is the violation of assumption 3 [17].

The null hypothesis in the ANOVA is that all group means are equal [15, pp. 309–362]. ANOVA returns an F-statistic or F-ratio telling whether the null hypothesis should be rejected or not. The F-ratio only tells if the group means differ from one another, i.e. whether the experimental manipulation has had any effect. Further tests are needed to find out how the group means differ. These tests can be either planned contrasts or post-hoc tests. Planned comparisons are used to test specific hypotheses, whereas post-hoc testing is more of an explorative approach used whenever there are no specific hypotheses about the data.

In post-hoc testing all different conditions are compared pairwise [15]. The problem in the comparison is that the probability of Type I error, i.e. the probability of finding false positive results in the analysis, increases as the amount of comparisons increases. To keep the error rate low enough a correction has to be applied. The significance level is typically corrected with the Bonferroni correction. The downside of the correction is the loss of statistical power, i.e. the probability of rejecting a genuine effect increases. Choosing the right type of correction is a trade-off between the Type I error rate and the statistical power.

In the ANOVA the observations are assumed to be independent (assumption 3). However, this is not the case when the same test subjects are used repeatedly over different conditions. A specific type of ANOVA allowing this kind of dependency in the data is called the repeated-measures ANOVA [15]. It makes an additional assumption about the data; it assumes that the variances of the differences between treatment conditions are roughly equal. This is called the assumption of sphericity. If the sphericity is violated a correction, such as the Greenhouse-Geisser correction, has to be applied.

2.3 Psychoacoustic phenomena

In the next sections the most essential psychoacoustic phenomena related to this thesis are presented. One particular aftereffect studied in this thesis, the Rosenblith aftereffect, is reported as an unusual "metallic" timbre in the sounds following the exposure. The section 2.3.1 introduces the concept of timbre, the theories associated with it and the scales that are used to measure it. The experiment of this thesis, explained in more detail in chapter 4, utilizes absolute hearing thresholds in the measurement of the effect. Different types of thresholds related to the hearing are discussed in chapter 2.3.2. It is important to distinguish between two types of absolute thresholds; detection of the sound in the first place and detection of AM or FM, i.e. temporal fluctuation, in the sound. The same holds for the masking phenomenon which is brought up in section 2.3.3; either a sound is masked by another sound or the modulation of the sound is masked by the modulation of another sound. Adaptation and fatigue, in section 2.3.4, are two central phenomena offering possible explanations to the experiment.

2.3.1 Timbre and modulation

When two sounds having the same pitch, loudness and duration, e.g. the same musical note on a piano and on a guitar, are played back to human listeners they can easily distinguish the sounds [6, p. 188]. They do it by listening to the "tone quality", also known as timbre, a property of sound which has no explicit definition. Timbre is a multidimensional psychoacoustic measure, which means there is no single scale with which to compare the timbres of two different sounds. The sensation of timbre is suggested to be related to the spectrum and its fluctuation over time [6, pp. 188–189]. In most cases, the sounds having a constant short-term spectrum have a constant timbre whereas changing the spectrum changes the timbre. However, also the onset of the sound can affect the perceived timbre.

The spectrum can fluctuate in two ways, either in amplitude or in frequency. The spectra of many everyday sounds, such as that of the human singing voice and many musical instruments, fluctuate either in amplitude, frequency or both [4, Chapter 7]. In engineering the fluctuation is called modulation, and the terms amplitude modulation (AM) and frequency modulation (FM) are used. Fig. 11 shows a graphical presentation of the basic principle of AM and FM of the tone (a), called the carrier signal, with the tone (b), called the modulator signal. Both the

AM and the FM can be presented mathematically. The carrier signal is denoted as

$$\sin(\omega_0 t), \tag{1}$$

where ω_0 is the angular frequency of the tone. In AM the carrier is multiplied with the modulator. This makes the carrier's envelope change according to the modulator. AM can be written as

$$p(t) = [1 + m \sin(\omega_m t)] \sin(\omega_0 t), \tag{2}$$

where $m \in [0...1]$ is the modulation index, also called the modulation depth, and ω_m is the modulation frequency. In FM the frequency of the carrier is modulated with the modulator which can be written as

$$p(t) = \sin[\omega_0 t + k \sin(\omega_m t) t], \tag{3}$$

where k is the width of modulation.

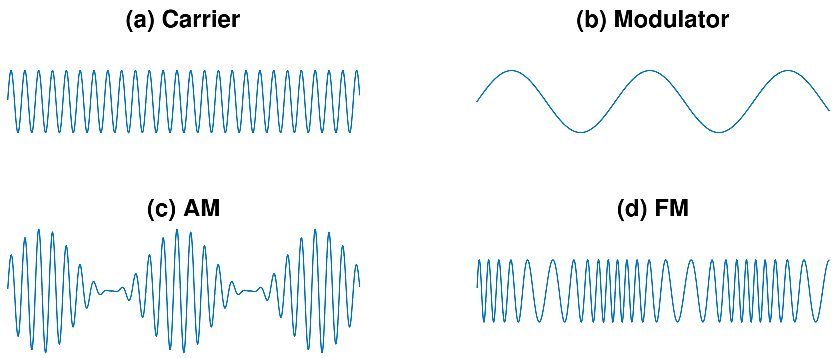

Figure 11: A basic example of modulations with pure tones. (a) The carrier signal $\sin(\omega_0 t)$, (b) the modulator signal $\sin(\omega_m t)$ (c) Amplitude Modulation, (d) Frequency Modulation.

When a sound is amplitude or frequency modulated with frequencies of about 1–16 Hz temporal fluctuation is perceived [12, chapter 10]. The lower limit of the fluctuation perception is due to the limitations in the short-term auditory memory and the upper limit due to the sluggishness of the hearing mechanisms. The strength of the fluctuation is measured in units called the vacil. The sensation of one vacil arises from 1 kHz sinusoid which is played at 60 dB SPL and has AM with modulation depth 100% and modulation frequency 4 Hz.

AM frequencies of about 15–300 Hz are not perceived as fluctuation but changes in a quality called the roughness instead [12]. The roughness starts to be audible already at the modulation frequencies 10–15 Hz. The unit of roughness is called the asper and the sensation of 1 asper is caused by a 1 kHz tone played at 60 dB SPL having amplitude modulation with the depth 100% and modulation frequency 70 Hz.

The detection of modulation in a sound signal depends on both the modulation frequency and the modulation depth. For amplitude modulation this relationship is illustrated by the temporal modulation transfer function (TMTF) in Fig. 12. The hearing system is most sensitive to modulations around 4 Hz, which also elicit the maximal sensation of fluctuation [12, Chapter 10]. This has been suggested to indicate a matching between the hearing and the temporal structure of speech since a fluent speaker pronounces about 4 syllables per second when speaking at a normal rate [18].

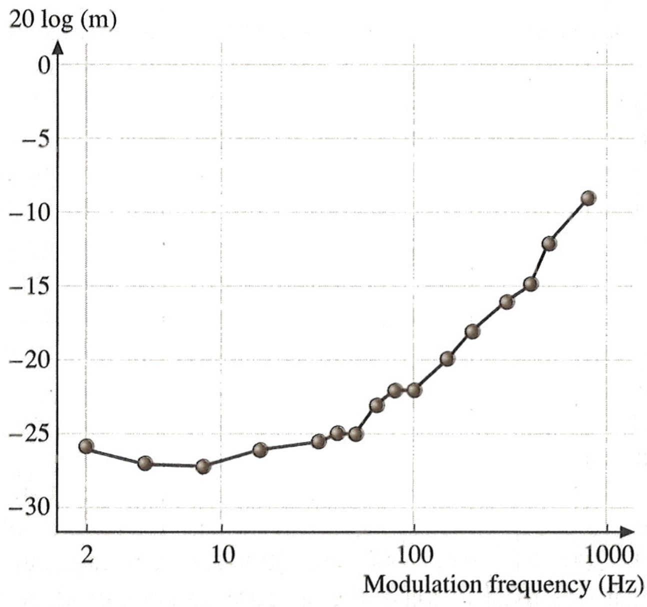

Figure 12: Temporal modulation transfer function (TMTF) after [19]. The threshold for detecting modulation in a sound signal depends on the modulation frequency and the modulation depth m, plotted in log-scale. Adopted from [7, p. 473].

2.3.2 Auditory thresholds

Two types of thresholds can be measured in the hearing system: absolute thresholds and difference thresholds [13]. The absolute threshold is the minimal sound intensity that a human can barely detect. The difference threshold is the minimal intensity difference between two sounds above the absolute threshold producing a perceptual difference.

In practice the absolute threshold varies due to the spontaneous activity and internal noise of the hearing system [13]. That is why the absolute threshold is defined as the intensity value that elicits perceived responses on 50% of the trials. Fig. 13 shows the absolute threshold of hearing in quiet measured with pure tones. What can be seen from the figure is that the auditory system is the most sensitive in the frequency range of 2–4 kHz due to ear canal amplification. The sensitivity develops during childhood and adolescence until the best sensitivity is achieved in the early adulthood [12]. The threshold shifts permanently up by age, especially at high frequencies, even in the absence of damaging noise exposure [4, Chapter 31].

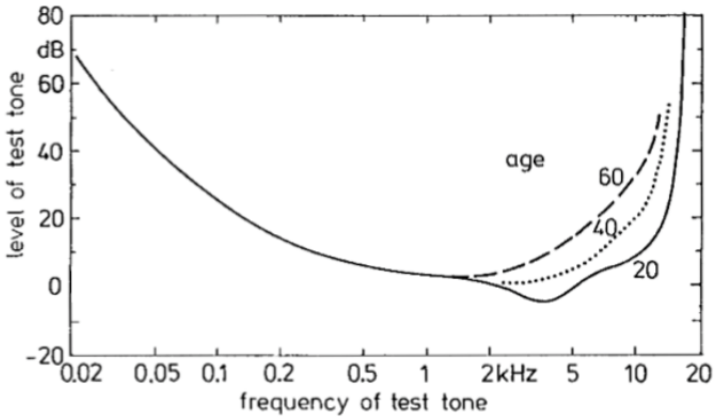

Figure 13: Average absolute thresholds in quiet for pure tones on the whole hearing range of different-aged people. The absolute threshold for 1 kHz is 0 dB. Adopted from [12].

The absolute threshold can also shift temporarily after exposure to loud noise [4, Chapter 31]. The magnitude of the temporary threshold shift (TTS) can vary from a few decibels lasting for a few hours to the ear being temporarily deaf and lasting two or three weeks. Typically the frequencies around 4000 Hz are affected the most. If the noise continues or repeats too frequently a permanent threshold shift can occur. This is due to the prolonged noise destroying the hair cells in the inner ear, more specifically in the organ of Corti. If the noise continues, eventually the organ itself is destroyed. The following are known about the relationship between noise and the TTS [20]:

1. Doubling the exposure time tends to double the threshold shift in dB.
2. Moderate TTS usually recovers within 16 hours. TTS above 50 dB may never recover completely.
3. Low-frequency noise produces less TTS than high-frequency noise.
4. TTS appear to be the greatest at a frequency one half to one octave higher than that of the fatiguing sound. Even if there is no TTS at the frequency of the fatiguing sound there can still be TTS of 15–20 dB at higher frequencies [21]. Increasing the intensity of the fatiguing sound moves the "center of balance" of the TTS pattern upwards in frequency [22].
5. An intermittent noise produces much less TTS than steady noise.
6. TTS appears to be entirely a physiological effect.

Hearing threshold can also shift temporarily due to a protective mechanism in the middle ear, known as the stapedius reflex or the acoustic reflex [4, Chapter 5]. When the sound pressure level reaches the level of 80–95 dB a set of muscles tightens the ossicles desensitizing the ear up to 20 dB. This protects the inner ear. However, it does not protect against impulse noise since it takes 30 or 40 ms for the reflex to begin and up to 200 ms for it to fully activate [4, Chapter 6]. The reflex takes place in both ears even if only one ear is stimulated [23]. The activation threshold depends on the bandwidth of the sound. For example, wide-band noise signals can activate the reflex even at sound pressure levels below 80 dB [24]. When the sound that activated the reflex is kept constant for a longer time adaptation in the admittance of the middle ear can be measured [23]. In other words, the reflex start to fade away, the rate of which is slower for low frequencies (e.g. 500 Hz) and higher for high frequencies (4000–6000 Hz). The rate of adaptation is measured in half-life, i.e. the time needed for the acoustic reflex to fade to half from the maximum level. For example, the half-time for broadband noise at 96 dB is about 60 seconds.

2.3.3 Masking

When two sounds that are close enough in frequency are played back at the same time only the louder sound can be heard [12, chapter 4]. In other words, the louder sound acts as a masker masking the softer one. The phenomenon originates from the overlapping excitation on the basilar membrane, shown in Fig. 14. Masking happens both in frequency and time, i.e. when two sounds are presented simultaneously and non-simultaneously. These phenomena are called spectral and temporal masking accordingly.

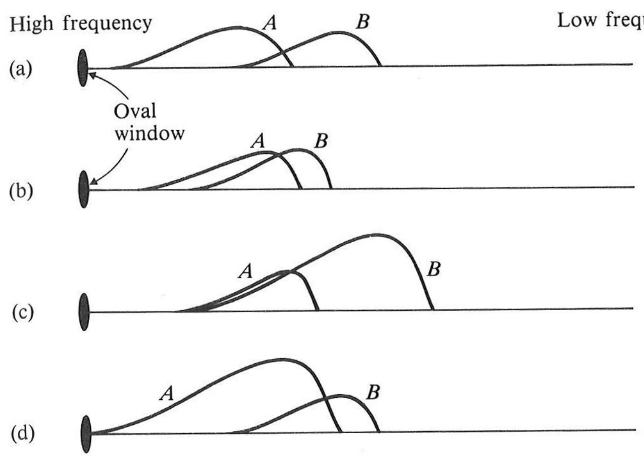

Figure 14: Excitation pattern of two pure tones on the basilar membrane. (a) The excitations barely overlap, thus not much masking occurs. (b) Tone B masks tone A because of the overlapping. (c) The more intense low frequency tone B almost completely masks the high frequency tone A. (d) The more intense high frequency tone A does not completely mask the low frequency tone B. Adopted from [4, p. 103].

Spectral masking means that the masker is the most effective in masking a softer sound if they have similar frequency content. The effect spreads asymmetrically in frequency, masking the neighboring higher frequencies more effectively than the lower frequencies. This, again, originates from the excitation pattern of the basilar membrane shown in Fig. 14. Temporal masking, shown in Fig. 15, can be divided into simultaneous, pre- and post-masking. The two latter ones are also referred to as backward and forward masking. The masker does not only mask the softer sound when they are presented simultaneously but also when the softer sound is presented before or after the louder sound. Pre-masking can occur for sounds presented 5-10 milliseconds before the masking sound. Post-masking typically lasts for 150 - 200 milliseconds, however, frequency content, amplitude and duration of the preceding sound affect the masking curve. Two explanations for post-masking have been proposed [25]. The first one is related to the ringing effect of the basilar membrane, in other words the basilar membrane keeps vibrating for some time after the off-set of a sound. The second explanation is related to the neural mechanisms that adapt to the masker.

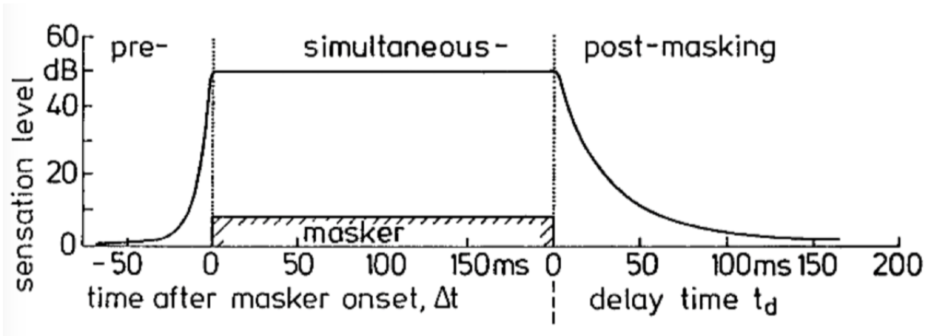

Figure 15: Pre- and post-masking by a sound called the "masker". Adopted from [12].

The masking effect also occurs in the modulation domain, in which case it is referred to as the modulation masking. For example, AM modulation in noise is harder to detect in the presence of another noise signal, the masker [26]. The detection becomes harder the closer the masker AM frequency is to the signal AM frequency and the greater the masker modulation depth is. The finding suggests that there are channels tuned to modulation frequencies the same way there are critical bands for the analysis of spectral content.

2.3.4 Auditory adaptation and fatigue

The neural responses in the auditory system decrease after prolonged exposure to a constant sound. In the context of hearing and this thesis the phenomenon is called adaptation. Adaptation explains the temporary loss of sensitivity, e.g. in the case of noise exposure. The decrease is not even in all neurons, instead it is the greatest in the neurons that are the most sensitive to the specific sound one was exposed to. In other words, feature-specific adaptation takes place. There is also another term, fatigue, referring to a phenomenon similar to adaptation. Although the two terms are oftentimes used interchangeably, there are five criteria by which they differ [27]:

1. Fatigue is typically associated with loud and long stimuli, whereas adaptation is usually measured after weak and short stimuli.
2. In adaptation the induced threshold shift does not cumulate during the first 0.05-10 seconds of stimulation, whereas in fatigue the threshold shift cumulates starting from 30 seconds up to at least 10 minutes.
3. In adaptation the plot of the absolute threshold returns back to the original value in a straight line whereas in fatigue the plot shows negative acceleration.
4. In adaptation the threshold shift occurs at the frequency of the stimulus whereas in fatigue the threshold shift might lie half octave higher.
5. Loudness recruitment, i.e. the increase of loudness with the sound intensity, differs in subjects showing either adaptation or fatigue.

2.4 Auditory aftereffects

Prolonged exposure to a constant or repeating stimulus can induce a transient alteration in the perception of subsequent stimuli [28]. This is commonly referred to as an aftereffect. Several aftereffects, or "afterimages", are known to exist in sight. One famous example is the waterfall effect. After watching a waterfall for a while fixed objects, such as rocks, look like they are moving [1]. While visual aftereffects are well known and have been widely studied, there is not that much knowledge nor studies about similar effects in the auditory modality [28]. This is because auditory aftereffecs are often elusive and require specific test conditions.

Several theories try to offer an explanation to the aftereffects. The theory of sensory persistence is related to the auditory memory and its capability of sustaining traces of the sound. In fact, traces of sound have been shown to persist as long as 10 seconds [29]. Another popular theory is related to neural adaptation in the auditory system. According to that prolonged exposure reduces the responsiveness of the neurons that are specifically activated by the features of the stimulus [28]. The decreased responsiveness in turn biases the perception of the subsequent stimuli causing them to appear different, e.g. to have an unusual timbre.

Studies on auditory aftereffects are sparse but they all typically approach the topic with the following questions:

1. How long does the effect last?
2. How strong is the effect?
3. What is the quality of the effect, i.e. how does it sound like?
4. How does the SPL of the inducer affect the magnitude of the effect?
5. How does the inducer duration affect the duration of the effect?
6. Does the effect transfer across the ears?
7. Does the effect occur in peripheral or central hearing?

Although the mechanisms behind the effects are still mostly unknown, the studies divide them roughly into effects occurring in the peripheral hearing and into those occurring in the central hearing. The underlying assumption is that effects taking place in the central hearing typically transfer across ears, last longer, and show stronger adaptation the higher in the hearing system the effect takes place [30]. In addition to the actual mechanism, it is unclear what kind of signal level characteristics a sound has to have in order to cause an aftereffect to emerge.

The following sections summarize several studies on auditory aftereffects. Section 2.4.1 reports how the DT for modulation can be elevated. Studies in which a sensitization in hearing was measured after exposure to loud low-frequency pure tones are summarized in section 2.4.2. Section 2.4.3 covers studies on short-term auditory afterimages, of which the most famous is the Zwicker tone. Finally, studies related to the metallic timbre lasting up to minutes induced by a specific kind of pulse train, the Rosenblith pulse train, are tackled in section 2.4.4.

2.4.1 Altered detection of modulation

In one study [31] a long exposure to a sinusoidally modulated 500 Hz pure tone at 50 dB SPL caused a loss of sensitivity in the detection of sinusoidal modulation. The study reported that FM always elevated the DT of FM more than the DT of AM. Similarly, AM elevated the DT of the AM always more than the DT of the FM tones. Both FM and AM were investigated at modulation rates 2, 4, 8, 16 and 32 Hz. The loss of sensitivity was induced slowly, requiring 20-30 minutes of exposure, whereas the recovery occurs within the first 60 seconds after the offset of the sound. However, it was not reported whether the subjects could actually hear the loss of sensitivity, for example, as a change in the timbre of everyday sounds.

The study [31] suggested separate channels for AM and FM in the hearing system – both having similar time course for adaptation and recovery. In another study [32] a prolonged exposure and training was found out to weaken the AM adaptation effect, making it completely disappear after about 10-12 hours in some cases. It is not known whether the same also takes place with FM.

2.4.2 Sensitization

When loud low-frequency tones were played for 3 minutes at 120 dB, after which the detection threshold was tested with clicks, an increase in the threshold was seen [33]. The threshold recovered back to the original after some minutes, however, in a non-monotonic way, shown in Fig. 16. The threshold continued to decrease for the first minute after the exposure, until after about two minutes the threshold rose again. This unexpected increase is called the Hirsh & Ward's "two-minute bounce", and is referred to as the bounce later on. The magnitude of the bounce reduces as the intensity of the inducing tone is decreased. Also, both the overall fatigue and the bounce are reduced in amplitude when the stimulation time is decreased. Several frequencies were tested and it was found that the different frequency regions of the auditory system recover at different rates.

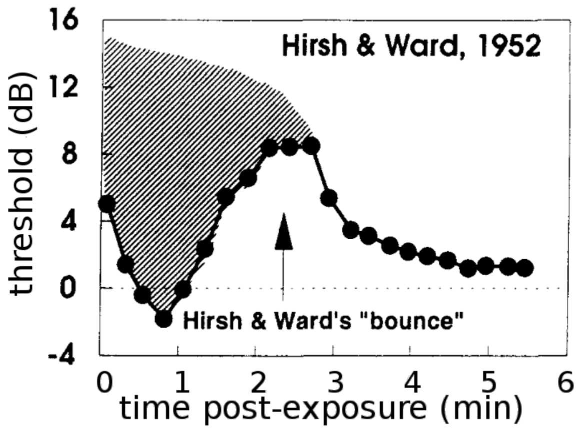

Figure 16: The Hirsh & Ward's "two-minute bounce" measured with psychophysical methods. The absolute threshold recovers in a non-monotonic way after exposure to loud low-frequency tones. At around 1 minute the hearing becomes more sensitive compared to the level before the exposure. Adopted from [34].

The subjects reported changes in the quality of tinnitus or head noise after exposure to the loud tones [33]. Also, the clicks following the exposure gained a definite thudlike quality. This was the case especially for the higher frequency tones. In particular, exposure to 4000 Hz pure tone at 120 dB did not only give the clicks a thudlike quality but also made speech sound distorted for up to one hour. However, the measured click thresholds were elevated only little or not at all.

A 500-Hz tone at 120 dB induced several effects [33]. Right after the exposure an unusually loud roaring noise was heard in the background continuing for about 70 to 80 seconds. After 200 seconds this noise assumed a "normal" level. Some subjects also reported hearing a 500-Hz tinnitus at the same time right after the exposure for about 20 seconds. When the clicks were played back, they were reported to "sound quite sharp at first but then gain the thudlike quality". This was concluded to mean that the low frequency components of the click were heard first. The change in the click quality happened coincidentally with the measured "two-minute bounce" which was concluded to indicate that the high frequency components become inaudible giving rise to the bounce, i.e. the raised threshold.

In typical psychoacoustic experiments the exposure to sounds causes fatigue and TTS, discussed in section 2.3.2. The elevated threshold in the beginning of the Fig. 16 is a clear sign of fatigue. However, after that the threshold starts decreasing. At around 1 minute the threshold level decreases even below the normal threshold, followed by the second rise, i.e. the two-minute bounce. In other words, the hearing becomes hypersensitive for a short moment. An otoacoustic study suggests that this sensitization is the oddity rather than the two-minute bounce, reported by Hirsh and Ward. The sensitization, which is induced by pures tones with frequencies below 2 kHz, is also known as the Kemp's "bounce" [34].

Auditory sensitization of as large as 8 dB was reported when measured 1 minute after exposure to 500 Hz tone at 112 dB for 3 minutes [35]. The same tone played even at 80 dB for 1 minute lowered the threshold for 1–2 dBs. The sensitization occured when low frequency pure tones were used as the exposure stimulus, whereas high and middle stimulation frequencies produced more fatigue than sensitization. A pure tone was found out to be able to sensitize the auditory system to a relatively wide range of test frequencies and, conversely, a relatively wide range of exposure frequencies sensitized the auditory system to the pure tone. This was considered consistent with what is known about the spread of energy, shown previously in Fig. 5. However, the study [35] does not report whether both the presence of the low frequency and the absence of the higher frequencies is required for the sensitization to take place. The study also found contralateral sensitization which was concluded to suggest that the location of the effect is in the central hearing system. Another possible explanation was considered to lie in the efferent pathways.

The mechanism of the sensitization or the Kemp's bounce has been studied in paralyzed guinea pigs [34] and cats [36]. In [34] all neural activity in the cochlea was blocked by using a neural paralyzing agent, tetrodotoxin. The sensitization was measured despite of this. Also, because the test animals were under paralysis the involvement of the middle-ear was ruled out as a potential mechanism. Thus, the sensitization was concluded to be related to non-neural cochlear mechanisms.

The basilar membrane velocity response to 500 Hz, the frequency studied both in [35] and [33], has been modeled [6, Chapter 7]. The response is shown at two different SPL levels in Fig. 17. The clearest difference between the two excitation patterns is that the excitation at the higher SPL spreads, especially to the higher neighboring characteristic frequencies. Another difference is that the basilar membrane reaches the maximum velocity sooner with the higher SPL.

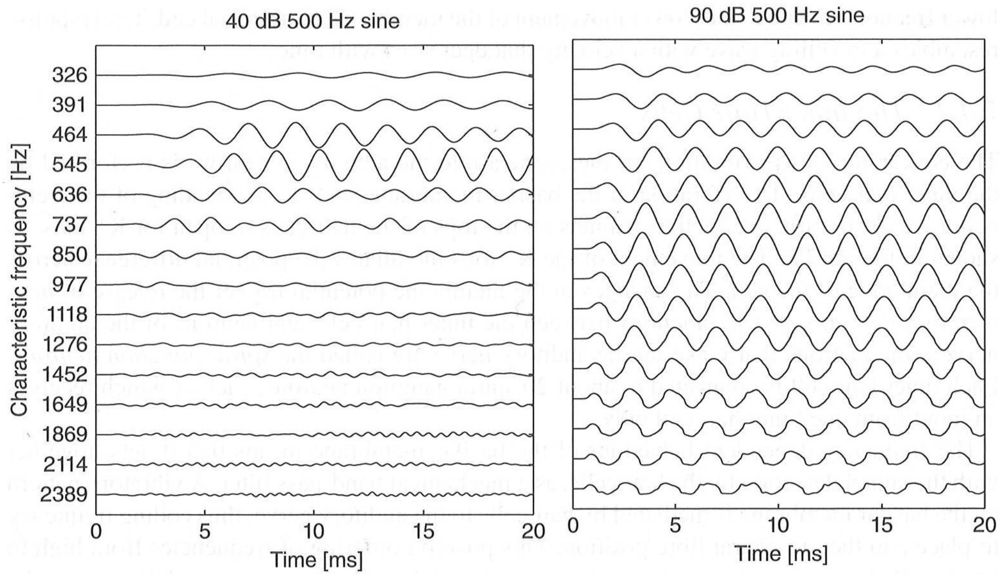

Figure 17: The normalized basilar membrane velocity response to 500 Hz pure tone at two different SPL levels. Courtesy of Alessandro Altoè. Adopted from [6, p. 121]

2.4.3 Auditory afterimage

When white noise, seen in Fig. 18, with a half-octave-band spectral notch between 300 and 7000 Hz is presented for 1 minute at 60 dB and switched off, a faint decaying tone can be heard for up to 10 seconds [2]. The audible tone, called the Zwicker tone, corresponds to the notch. In other words, the heard tone is as if the "negative afterimage" of the noise spectrum. At very low and very high sound pressure levels the Zwicker tone does not appear at all. The effect does not transfer between ears.

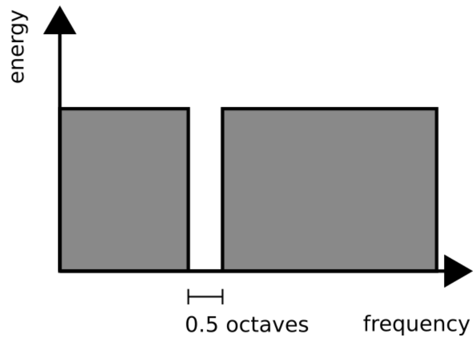

Figure 18: White noise with a spectral notch induces an illusion of a tone called the Zwicker tone.

Despite the fact that the Zwicker tone is probably the most studied of all the auditory aftereffects, the mechanism creating the effect is not known. The tone cannot be explained by the known mechanisms in the peripheral hearing alone [37]. Some studies have suggested potential neural correlates at the auditory cortex both in humans [38] and cats [39]. It has also been suggested that the mechanism behind the Zwicker tone and tinnitus are similar [40].

A similar afterimage effect that, however, vanishes much faster has also been demonstrated with a more complex sound. When the complemented spectrum of a vowel is played back for more than 150 ms, followed by a quick transition to uniform spectrum, the original vowel is heard for up to 500 ms [41]. The stimulus is illustrated in Fig. 19. A similar auditory afterimage has been reported while listening to broadband noise followed by sinusoidally comb-filtered noise [42]. Another example is the experiment in which a harmonic series with a missing component is played back [43]. When the missing component is reintroduced it is heard clearly on top of the existing harmonics. The auditory afterimage does not transfer across the ears in any of the abovementioned experiments. Since the effect also fades away quickly, in less than 500 ms in [41] and in 3-6 seconds in [42] and [43], it has been suggested to be related to peripheral adaptation processes.

2.4.4 Metallic timbre

The timbre of environmental sounds, such as a handclap, the sound of rubbing sandpaper or own voice, changes to "metallic" for a few seconds after exposure to a pulse train of about 100 Hz, seen in Fig. 20 [3]. This pulse-train effect, referred to as the "Rosenblith aftereffect" in this thesis, can last from a few seconds [3] up to a

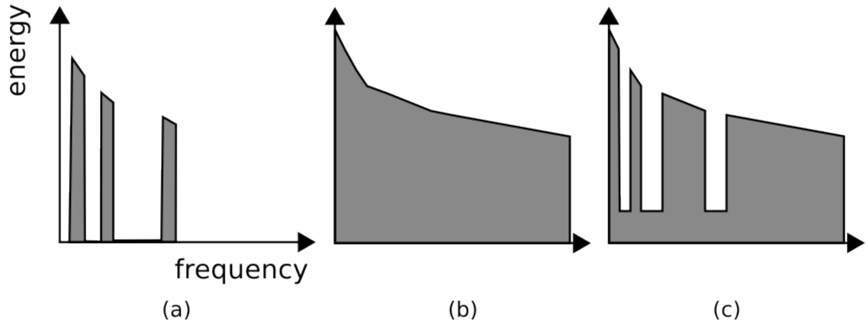

Figure 19: When the spectrum of a vowel (a) is subtracted from the uniform spectrum (b) the inverted spectrum (c) is obtained. Subjects who listen to the inverted spectrum followed by a quick transition to the uniform spectrum report that they hear the vowel plotted in (a).

few minutes [28]. Several parameters of the pulse train were tested in the original experiment [3]. The pulse train frequencies between 30 and 200 pulses per second were found to be the most effective in producing the aftereffect. By testing inducer durations from 5 to 240 seconds the longer exposure was found to elicit a longer effect. It was also concluded that the more intense sounds prolong the effect. Training did not change the effect duration, nor did it transfer across ears. In addition to the pulse train, other types of stimuli were tested. A square wave was found to elicit the effect, however not being as effective as the pulse train. Noise pulses or a non-harmonic tone repeated at the same rate as the pulse train, however, did not elicit the effect. High-frequency components in the sound were found to be necessary for the phenomenon. For some listeners very loud pure tones also made the Rosenblith effect occur.

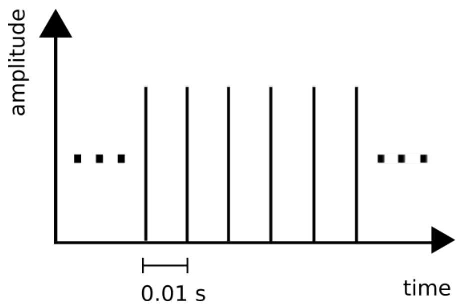

Figure 20: Rosenblith pulse train at 100 Hz.

The Rosenblith effect was revisited by [28] and new kinds of subjective effects were reported. It was discovered that the pulse train serves as its own test stimulus. When listening to the pulse train for more than 1 minute the roughness that was there in the beginning faded away over time. At the same time the buzzing element of the sound became more prominent and segregated. However, the main finding of the study [28] is related to the detection of modulation. The DT of AM frequencies between 100 and 500 Hz was found to elevate for 30 seconds after listening to a 100

Hz pulse train at 46 dB SPL for 60 seconds. The effect did not transfer across ears. The Rosenblith effect was concluded to be related to a temporary alteration in the perception of fast temporal envelope fluctuations of about 100–500 Hz. It was argued that this alteration cannot be explained by adaptation to AM, as was the case with modulated pure tones presented in section 2.4.1. This is because of three differences [28]: the tested modulation frequencies were much higher, the effect did not transfer across ears and the adaptation persisted despite considerable exposure and training.

The exact mechanism causing the Rosenblith aftereffect is currently still unknown [28]. Because the effect has not been reported to transfer across ears it is highly likely occurring in the peripheral hearing. Thus, studying the movement of the basilar membrane during the pulse train might provide hints about the mechanism. Fig. 21 shows the modeled movement of the membrane during the continuous excitation by the 100 Hz pulse train. The figure clearly shows how the lower characteristic frequencies are excited with a delay to the higher frequencies. At higher frequencies the excitation pulsates according to the pulse train rate, whereas at lower frequencies the basilar membrane is in constant movement.

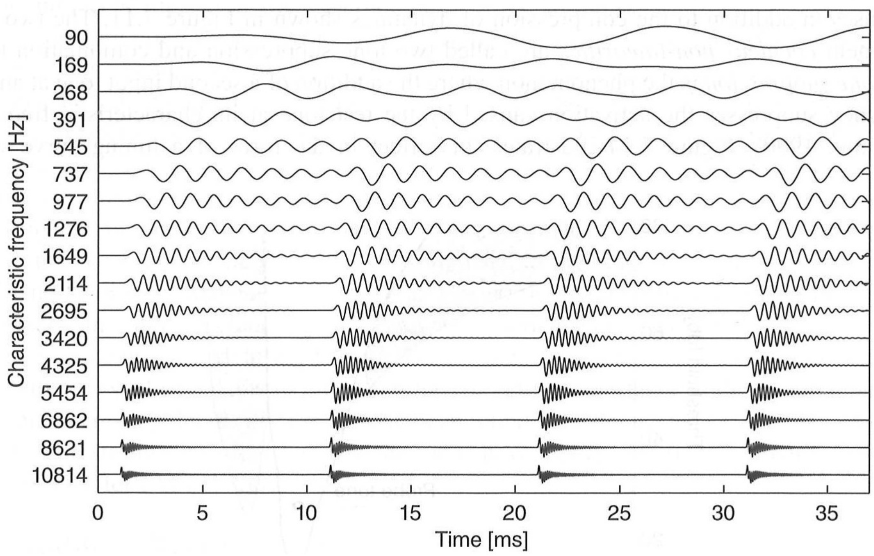

Figure 21: The velocity response of the basilar membrane to the impulse train with a rate of 100 Hz, based on the model reported in [44]. Adopted from [6, p. 121]. Courtesy of Alessandro Altoè.

This chapter provided an overview of the sparse studies on auditory aftereffect. The introduced effects included both altered DTs (sections 2.4.1 and 2.4.2), illusions of sounds (section 2.4.3 and changes in timbre (section 2.4.4). The effect durations varied from hundreds of milliseconds in the vowel effect, up to the minute scale observed in the sensitization. All effects became more intense and lasted longer when the exposure time and the exposure SPL were increased. None of the effects, expect FM adaptation, transferred across ears. This indicated the mechanisms to lie in the peripheral hearing. Despite of the presented studies, the nature and the cause of the aftereffects remain unknown.

3 Noise detection thresholds after exposure to pulse trains

As previously mentioned, aftereffects are still poorly understood. The experiment conducted in this thesis aims at finding out whether perceptual aftereffects, i.e. changes in perception after exposure, are associated with changes in hearing sensitivity. The study is conducted by analyzing the detection thresholds (DTs) after exposure to sounds that either induce or do not induce the pulse-train effect, referred to as the Rosenblith aftereffect (section 2.4.4). If the hearing sensitivity changes under the influence of the aftereffect, there should be a measurable changes in the DT. Test stimuli with similar signal characteristics as the Rosenblith pulse train were generated for the experiment. The Rosenblith effect was chosen as the starting point because it typically lasts rather long, unlike the aftereffect related to the auditory afterimage (section 2.4.3). In addition, an exposure at a comfortable sound pressure level is sufficient, unlike in the effect caused by the loud low-frequency pure tones (section 2.4.2).

The stimuli included sounds known to elicit the aftereffect [3]: the original 100 Hz Rosenblith pulse train (R) and its frequency modulated version (Rf). In addition there were two sounds known not to cause the aftereffect: noise (N) and harmonic noise (H). More detailed information about the stimuli can be found in section 4.2.1. In the experiment the stimuli were grouped in conditions in which they were used as a 60 second inducer at 65 dBA or as a test tone for which the DT was measured.

3.1 Research questions and hypothesis

This research had three goals which are approached by measuring the detection thresholds with psychophysical methods:

1. Is the DT different after exposure to an aftereffect-eliciting sound (R, Rf) compared to an aftereffect-free sound (N, H)? In particular, is there a TTS associated with the aftereffect?
2. Does listening to a specific sound alter the DT of that sound only? For example, is it easier or harder to detect R than some other sound after listening to R for a long time?
3. How does the time course of the DTs behave after exposure to the stimuli?

The experiment was divided into eight conditions. There were two conditions per stimulus: one where the inducer and the test sound are the same and one where the test sound is N instead. The latter served as control conditions. N was tested against itself and against the H. The tested conditions are listed in Table 1. Next, the research questions are reviewed based on what is currently known about the aftereffects and hearing in general. Hypotheses concerning the experiment results are also made.

Is the DT different after exposure to an aftereffect-eliciting sound compared to an aftereffect-free sound? As it was discussed in section 2.3.2,

Table 1: The experiment had eight different conditions with different inducer and test sounds.

Inducer	**Test sound**	**Name**
noise	noise	NN
noise	harmonic noise	NH
Rosenblith	noise	RN
Rosenblith	Rosenblith	RR
harmonic noise	noise	HN
harmonic noise	harmonic noise	HH
FM Rosenblith	noise	RfN
FM Rosenblith	FM Rosenblith	RfRf

listening to steady loud noise (N) is generally known to cause TTS and thus also to elevate the DT of the subsequent test sounds. However, TTS is not caused by loud noise only – even 20 dB pure tones have been reported to cause a measurable TTS [21]. Thus, all stimuli used in the experiment are expected to induce TTS the (absolute) amount of which, however, is out of scope of this study.

Instead of the changes in absolute DTs, the focus of the study is in the DT differences. This is why no conditions containing only the test sound were included. Regarding the differences between conditions, similar signals known not to cause aftereffects (N, H) are expected to behave the same way in terms of the induced TTS and the DT. To be more specific, no significant differences are expected in the DTs of conditions NN, NH, HH and HN. In contrast, it is not known whether the DTs after the stimuli eliciting the Rosenblith aftereffect (R, Rf) are different from stimuli not eliciting the aftereffect (N, H).

Does listening to a specific sound alter the DT of that sound only? According to the theory of adaptation, introduced in section 2.3.4, listening to a steady sound for a long time decreases the responsiveness of the neurons detecting the specific characteristics of the signal. If the exposure continues long enough even fatigue occurs. In the conducted experiment this means that using the same sound both as the 60 second long inducer and the test sound should show an elevated threshold compared to using different sounds. More specifically, the DT of the test sound N is expected to be higher in NN than in RN, HN or RfN. In the same way the threshold in HH is expected to be higher than in NH.

How does the time course of the DTs behave after exposure to the stimuli? There are several time scales related to the changes in DTs. The perceptual Rosenblith aftereffect typically lasts for a few seconds [3], during which changes in DT could occur. However, partly limited by the chosen method, the focus of the study is in longer-term changes in DT happening on the scale of tens of seconds to minutes. All stimuli are expected to elevate the DTs for up to several minutes. The magnitude and the overall shape of the DT recovery curve are unknown. However, the shape is eventually expected to show exponential decay, as is the case in Fig. 16

and generally with noise-induced TTS [4, Chapter 31].

Theoretically R and Rf could induce sensitization, as seen with loud low-frequency pure tones (section 2.4.2), occurring as a bounce in the DT within two minutes after the exposure. There are several reasons to assume this. Both the Rosenblith aftereffect [28] and the sensitization [34] are assumed to be related to peripheral mechanisms, the processes taking place in the cochlea. In fact, when the basilar membrane excitation patterns of the 100 Hz pulse train and the 500 Hz pure tones are compared (Fig. 17, Fig. 21), many similarities can be found. First, both create a continuous excitation on several low characteristic frequencies that do not change over time. In contrast, a soft 500 Hz tone would excite only a single characteristic frequency. Second, both R and Rf excite high frequencies either non-continuously in a pulsating manner or not at all. In contrast, N and H continuously excite all characteristic frequencies. The third similarity is related to the signal characteristics. When the R/Rf pulse trains and the pure tone are normalized to have the same RMS value, the peak dB value of the pure tone is about 50 dB lower. Thus, in order to transmit the same peak sound pressure to the inner ear the pure tone has to be played back at a much higher SPL measured in RMS. In fact, starting from SPLs of 80-95 dB the level has to be even more than 50 dB higher to overcome the attenuation caused by the acoustic reflex in the middle ear. However, if the sound is kept constant for a longer time the acoustic reflex starts to fade away [23]. Overall, although it is still an open question whether the studies conducted with loud pure tones can be applied to pulse trains, let alone generalized to broad-band signals having a more complex structure, the abovementioned similarities in the basilar membrane excitation patterns indicate the possibility for R to induce sensitization.

If R and Rf were to induce sensitization, it would be seen as a bounce in the DT after two minutes of exposure. However, possible sensitization by R would not necessary have a connection with the actual perceptual effect, the "metallic" timbre. This is because many studies mention that the high frequency componant are necessary for perceptual aftereffects, e.g. the Rosenblith effect [3] or distorted speech after 4000 Hz [33]. The sensitization, on the other hand, has only been observed for low frequencies below 2 kHz [34]. Thus, sensitization and the metallic timbre should be handled as two separate phenomena.

3.2 The chosen method

There are several requirements for a good measurement method that grasps the Rosenblith aftereffect by measuring the DT. The time window for the measurement is limited from 200 ms to some seconds after the offset of the aftereffect-eliciting sound. The lower limit is due to post-masking, introduced in section 2.3.3, that makes the thresholds rise for 200 ms [45]. The upper limit is due to the effect duration, reported in [3]. Although the time window is short, it is still longer than the shortest aftereffects, e.g. the auditory afterimage that can fade away in just 500 ms [41].

It is known that the perceptual Rosenblith aftereffect fades out in a few seconds [3]. In order to produce consistent results the method has to keep the effect "on" during the measurement procedure by repeating the inducer in appropriate intervals.

The measurement should also be kept as short as possible because of at least three reasons. First, the sound was considered annoying by many subjects. Second, the aftereffect was known to last for a rather short time. Third, the subjects generally find it hard to concentrate in longer experiments, which again adds noise to the data.

The abovementioned requirements were fulfilled by combining several known methods. An adapted version of the method presented in [28] was found to be a suitable basis for the experiment. The transformed up-down method, discussed in section 2.2.3, was used for adaptive tracking. It was found to be a relatively good compromise between the accuracy of the results and the test duration. The method of constant stimuli, despite of its simpler design, was not chosen since it would have made the experiment much longer. One particularity in the experimental design was the attempt to keep the aftereffect on by repeating shorter versions of the inducer sound, called the refresher sounds, between test stimuli. However, there were no previous studies that would have confirmed the effectiveness of these sounds. More information about the methods used in the experiment can be found in section 4.2.2.

4 Experiment

A listening test was conducted to investigate the hypotheses. All the subjects participated in an audiometric screening prior to the actual experiment. The results of the audiometry and the experiment are presented in sections 4.3 and 4.4. The description of the experimental setup for both can be found in section 4.2.

4.1 Test subjects

14 unpaid subjects (12 male, 2 female, age: 24–33, average age 26.93) participated in the listening test voluntarily. All subjects were staff of the Department of Signal Processing and Acoustics at Aalto University or students of the university. All were naive with respect to the purpose of the study. The author did not participate. The study protocol was approved by the Aalto University Research Ethics Committee. In this experiment normal hearing is defined as:

1. pure-tone thresholds below 20 dB HL in both ears at all tested frequencies: 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 8000 Hz
2. average over all the tested frequencies and both ears below 15 dB HL
3. maximum difference of 10 dB HL between the left and the right ears at any frequency

By the abovementioned criteria 11 subjects had normal hearing. Three subjects had differences greater than 10 dB between the two ears at single frequencies, but the average was still within normal limits. For subjects #2 and #15 the frequency was 1000 Hz and for subject #16 the frequency was 4000 Hz. These subjects were still included in the analysis because the results obtained from the actual listening test did not show deviation from other subjects.

4.2 Experimental setup

To assess the hearing of the subjects an audiometry was performed manually with a calibrated Interacoustics AD28 diagnostic audiometry test and closed-back Telephonics TDH-39P headphones. The audiometry took place in a listening booth where the noise floor was below the hearing threshold at all frequencies. The frequencies were tested in the following order: right ear 1000 Hz, 500 Hz, 250 Hz, 2000 Hz, 4000 Hz, 8000 Hz; left ear 1000 Hz, 500 Hz, 250 Hz, 2000 Hz, 4000 Hz, 8000 Hz. Each frequency was tested until a clear converge took place. The testing took about 20 minutes per ear. The subjects had a few-minute pause outside the listening booth between the tested ears. During the pause the listening booth was ventilated.

A schematic drawing of the setup in the actual experiment is shown in Fig. 22. The experiment was conducted in an anechoic chamber which had a noise floor below the hearing threshold at all audible frequencies. A computer running the Max MSP 7.0.3 software outside the chamber was used to present the sounds and to collect the responses. The responses were given with an iPad over WiFi. The sounds were played back through RME Digital / Analog Interface M-32 DA sound card and Sound Devices HX-3 headphone amplifier at a 48-kHz sampling rate. Open-back Sennheiser HD-650 headphones were used to present the same sound to both ears, diotically.

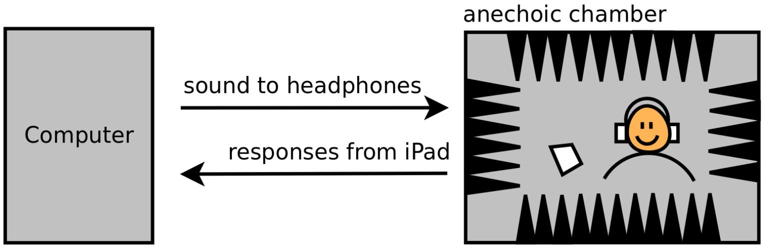

Figure 22: In the actual experiment the sounds were played through headphones to a subject sitting in an anechoic chamber. The responses were given with an iPad.

4.2.1 Stimuli

Four sounds, shown in Fig. 23, were used as the stimuli. Their names and abbreviations are: noise (N), harmonic noise (H), Rosenblith pulse train (R) and frequency modulated Rosenblith pulse train (Rf). Only two of the stimuli, R and Rf, were informally found to have an aftereffect, i.e. alter the perception of sounds presented after them.

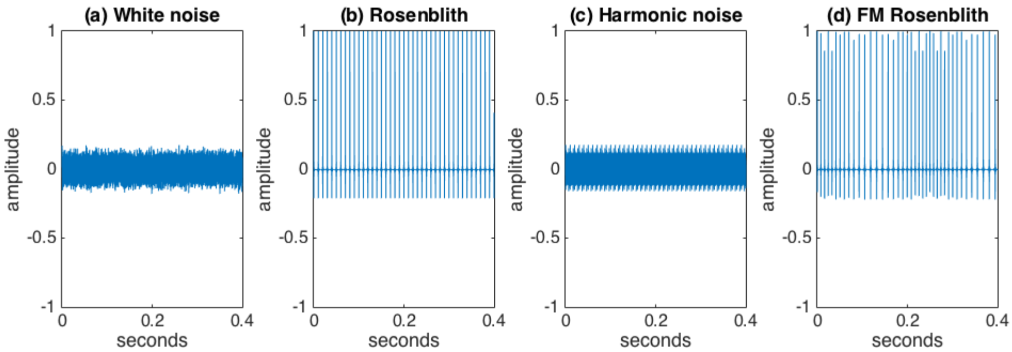

Figure 23: 0.4 seconds of the stimuli used in the experiment. N is white noise filtered to 100–16000 Hz. H is a sum of 100 Hz sinusoids and its harmonics up to 16 kHz. R is a pulse train with 100 Hz rate. Rf is the same as R but with 4 Hz frequency modulation. All the stimuli have the same RMS value.

The stimuli have some differences. All, except N, have a periodic structure. Also, compared to the R and Rf (0 dB) the peak values of N and H are about -30.6 dB and -37.4 dB lower. Despite of these differences all stimuli fulfill the following requirements:

1. has a flat long-term spectrum
2. contains only frequencies from 100 to 16000 Hz

3. has the same RMS, i.e. contains the same amount of acoustic energy, as all the other stimuli

N was generated by band-pass filtering white noise with a 5th order butterworth IIR filter. Stimuli R, Rf and H were generated by summing 100 Hz cosine wave and its harmonics up to 16 kHz, thus they naturally fulfill the requirement 2. H has a similar spectral structure as R, the only difference being in the phases of the summed cosine waves: in H the summed cosine waves are in random phase, in R they are in the same phase. Rf was obtained by frequency modulating R with the modulation frequency 4 Hz so that the pulse rate varied approximately between 75 and 125 Hz. Modulation frequency 4 Hz was chosen because the hearing is known to be the most sensitive to it [12]. In order to fulfill the requirement 3 all the sounds were normalized. The maximum SPL at which the stimuli were presented was 65 dBA, calibrated with B&K Type 2250 Hand-held Analyzer and B&K Type 4189 microphone capsule. For more detailed description of how the stimuli were generated see the Matlab code in Appendix A.

4.2.2 Procedure

A method adapted from [28] was used to investigate the effect of the stimuli on DTs. However, to make the effect more intense some experiment parameters were adjusted. The inducer was presented louder than 46 dB SPL, at 65 dBA. Also, the refresher duration of 10 s was chosen instead of 4 s. The changes were made because an intense and long exposure is known to induce the aftereffect the most effectively ([3], [33]).

The experiment had eight conditions in total. The timeline for one condition of the experiment is shown in Fig. 24. The timeline consisted of three parts called the inducer, the Alternative Forced Choice (AFC), and the refresher. The inducer appeared only once, in the beginning the condition. The AFC and the refresher, however, were repeated until a certain convergence criterion or maximally 50 repetitions were met. One repetition or a trial, called a "step" (not to be confused with dB steps), lasted between 12.1 and 12.7 seconds, on average 12.4 seconds. The step durations were uniformly distributed.

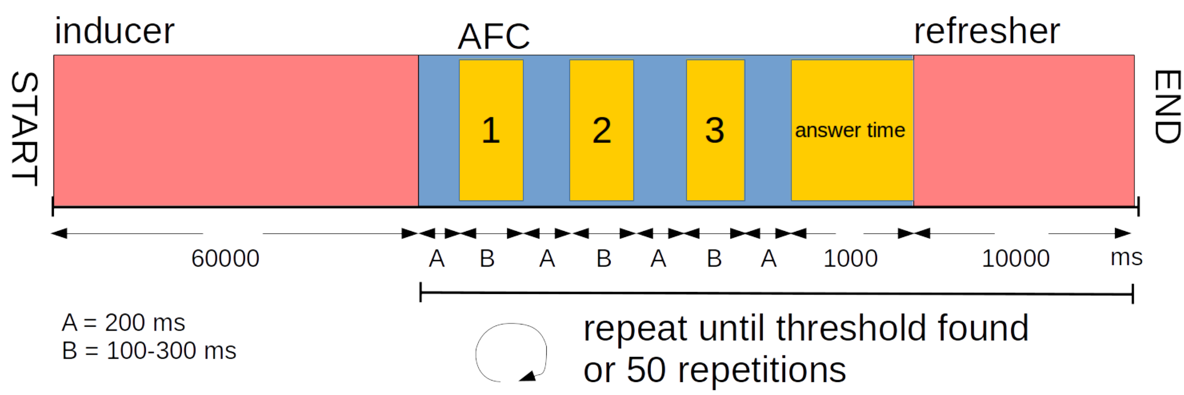

Figure 24: The timeline of one condition.

The inducer sound, a 60 second sound presented at 65 dBA, was followed by a pause of 200 ms. This was to prevent post-masking from interfering with the next stage, the AFC. In the AFC three buttons, separated by pauses of 200 ms, were flashed to the subject one at a time. In order to minimize subject's anticipations the flash durations were uniformly random between 100 and 300 ms. A sound called the "test sound", with the duration of 100–300 ms, was played back at the same time as one of the buttons is shown. The test sounds were randomized and uniformly distributed among all buttons. Next, the subject was given the time of 1000 ms to choose the button during which he or she heard the test sound. The response was saved and used to control the SPL of the test sound in the future AFC repetitions. The subjects were instructed to guess in unsure situations. This was because according to the signal detection theory subjects tend to answer correctly around the threshold even if they are not aware of it themselves [13]. If the subject did not answer anything for some reason, the test sound level was not left unchanged for the next step.

Regardless of whether the subject responded or not, the condition continued with the repetition of the inducer sound. However, this time the duration was only 10 s, which is why the sound is called the "refresher" rather than the inducer. In order to prevent audible clicks, produced by transients before and after the stimuli, all stimuli were smoothly faded in and out. The inducer and the refresher sounds had fade-in and fade-out ramps of 200 ms, whereas the test sound had fade-in and fade-out ramps of 20 ms. The inducer sound stimulus was played more than 80% of the duration of the condition.

The order of the conditions was counterbalanced and the subject had 1-2 minute pauses outside the chamber between the presentation of the conditions. The purpose of the pauses was to prevent the potential aftereffect from the previous condition from affecting the next one.

The procedure aimed at finding the DT of the test sound with an adaptive method, i.e. by adjusting the test sound level according to the responses. In the original method [28] the step size was initially set to 4 dB, reduced to 2 dB after the second reversal and to 1 dB after the fourth reversal. However, in this research the transformed up-down method, described in section 2.2.3, was used instead. Since the subject-specific DTs for the stimuli were not known beforehand the test sound level had to be set as high as 20 dBA SPL for the first AFC. To speed up the convergence, the level was lowered in 3 dB steps until two turning points were found. After that the steps were reduced to 1 dB. The end rule was defined as "six consecutive turning points within a 4 dB range". An example of this is shown in Fig. 25.

The interface presented to the subject is shown in Fig. 26. It has a progress bar on top showing the overall progress of the experiment. The vertical progress bar on the left shows the time left for this part of the condition, i.e. the inducer, refresher and the answer time left. To keep the subject attentive and to make the experiment more comfortable the subject was given feedback on his answer by flashing text "correct" or "wrong" immediately after the response. The subjects were instructed to press the "sound on/off" button, shown on the top in the interface, and come out of the anechoic chamber if they felt uncomfortable because of the stimulus. However,

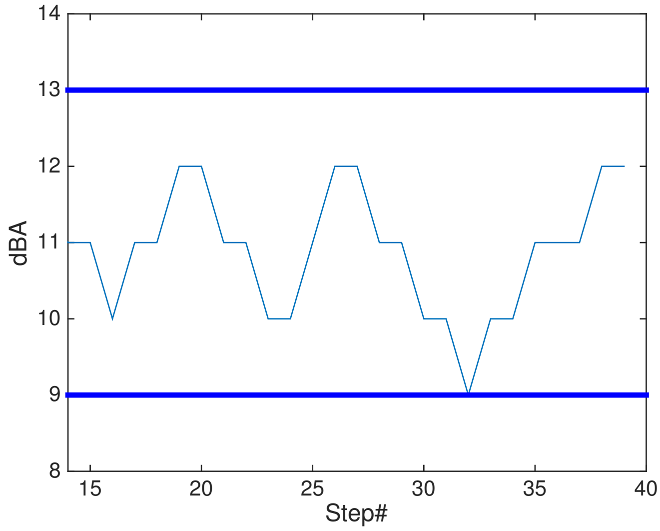

Figure 25: The track was considered converged when six consecutive turning points within a 4 dB range were found.

no subject pressed the button. The subject had short familiarization trials before the actual test to learn the interface.

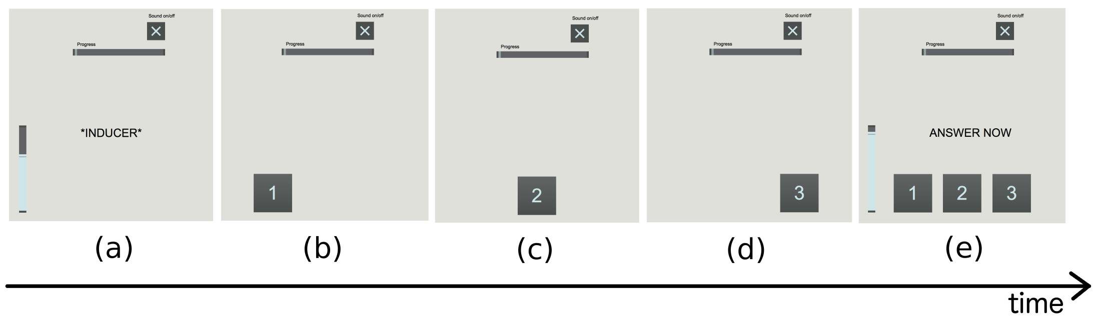

Figure 26: The interface presented to the subjects. (a) The inducer or the refresher sound playing, (b) button 1 flashing, (c) button 2 flashing, (d) button 3 flashing, (e) answer time. The test sound was played at the same with the button 1, 2 or 3. The subject got feedback in a text form ("correct" or "wrong") after responding.

4.3 Audiometry results

The averaged audiometry results over all the 14 subjects plotted with error bars are shown in Fig. 27. The highest of the tested frequencies, 4000 and 8000 Hz, show the greatest deviation between the individuals. The most important finding visible in Fig. 27 is how the audiograms differ per ear. Namely, the results for the right ear are notably worse at frequencies 500 and 1000 Hz. The most potential explanation for the difference lies in the order in which the ears and the frequencies were tested: the first three tested frequencies in the whole audiometry were 1000 Hz, 500 Hz and 250 Hz in the right ear. Thus, it is highly likely that the differences are due to inexperience of the subjects. Familiarization trials and counterbalancing could help to prevent this bias.

In total 14 subjects were qualified for the actual experiment. However, more potential subjects were screened. The reasons for discarding subjects in the audiometry were an unusually high threshold at an individual frequency (N=2), most often 4000 and 8000 Hz, a difference too great between the ears at some particular frequency or both (N=4).

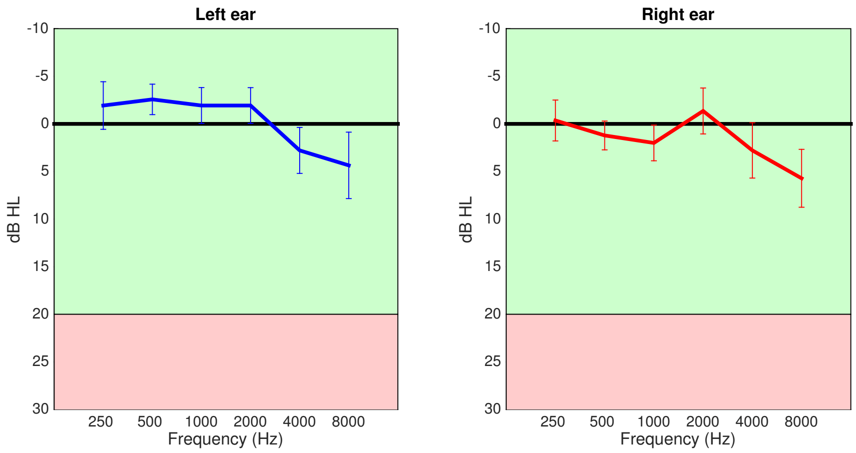

Figure 27: Average audiometry over 14 subjects for both ears. The error bar denotes the standard error.

4.4 Experiment results

The adaptive tracks obtained from the 14 subjects were divided into six different data sets, listed in Table 2, in order to find possible time-varying effects. The stepwise data sets contain the dB-values of the adaptive tracks on a trial-by-trial basis. The step ranges of the sets were defined by the shortest track which was only 19 steps long. This also means that the steps 1-50 have unequal sample size starting from step 20. For a more detailed presentation of the obtained tracks see the Appendixes B1 and C1, where the tracks are plotted per condition and per subject.

The data set "Final thr." contains threshold values calculated from last four turning points of the converged tracks. There are 112 threshold values in total, one for each condition and subject. 13 of the 112 recorded tracks did not converge to the range but had enough turning points to calculate the final threshold. A Shapiro-Wilk normality test was performed for the final thresholds per condition. The results along with the mean values are shown in Table 3. The p-values in conditions NN and HH are smaller than 0.05, thus they are not normally distributed. In the rest of the cases the normality cannot be excluded.

Table 2: The six analysed data sets, given with the amount of data points and the coverage of the track in minutes.

Data set name	Data points	Time after the inducer
Final thr.	112	-
Steps 1-50	3881	~ 0-10.1 min.
Steps 1-19	2128	~ 0-3.7 min.
Steps 1-10	1120	~ 0-1.9 min.
Steps 5-15	1120	~ 0.8-2.9 min.
Steps 9-19	1120	~ 1.7-3.7 min.

The subjective quality of the aftereffect or the stimuli used in the experiment were not analyzed. However, the subjects informally reported the inducers to have different loudnesses: N being the softest, R and Rf being the loudest. Also, R and Rf were reported to be the most annoying whereas N was considered the least annoying. The annoyance of H was between these two extremes. Some subjects did not notice a difference between N, R and H when they were used as test sounds, i.e. presented around the DT.

The results of the experiment are presented in the following sections. The analysis of button presses, track lengths and the response time course can be found in sections 4.4.1, 4.4.2 and 4.4.3 accordingly. The correlation between the final thresholds and the audiometry results are tested in section 4.4.4. Statistical analysis using ANOVA is conducted in section 4.4.5.

Table 3: The results of the Shapiro-Wilk normality tests and the mean values of the final thresholds per condition. The significant results are highlighted.

Condition	Mean	Shapiro-Wilk W	p-value
NN	11.61	0.866	**0.037**
NH	11.95	0.939	0.404
RN	12.45	0.925	0.260
RR	11.13	0.965	0.802
HN	11.93	0.958	0.683
HH	12.02	0.867	**0.039**
RfN	12.57	0.913	0.175
RfRf	12.13	0.896	0.099

4.4.1 Button presses

The response behavior of the subjects was analyzed in order to identify anomalies in the behavior and biases induced by the experimental setup. The button presses in the AFC part of the experiment were recorded as being "correct", "wrong" or having "no_answer". During the whole experiment there were only few situations in which the subject gave no answer. The collected data was used to calculate the hit rate, i.e. the ratio of the correct presses, for each button. These are reported in Table 4. Because of uniformly distributing the test sounds among the buttons in the experimental design, the hit rates were assumed to be similar with each other.

Table 4: Button hit rates in the whole experiment.

Button #	**Hit rate**	**Presented after the inducer sound (ms)**
1	0.7329	200
2	0.8403	500-700
3	0.7228	800-1200

4.4.2 Track lengths

The convergence rate of the adaptive tracks was examined by analyzing the track lengths. The key idea was that a longer track indicates slower converge, which in turn is a sign of either achieving a lower threshold or, more importantly, temporal fluctuation in the DT during the measurement. The fluctuation in turn could be a consequence of increased sensory noise due to the preceding exposure.

Fig. 28 shows the mean track lengths per condition and per subject. The condition producing on average the longest tracks, 38.29 steps (8.6 minutes), was the RN, whereas the shortest track, 30.64 steps (6.9 minutes), was produced by HN. On average the subject #6 took the longest time per condition, 41.13 steps (9.2 minutes), whereas the subject #14 needed only 28.75 steps (6.4 minutes). The tracks showing the slowest converge were RN and HH. The fastest convergence took place in tracks HN, RfRf and RR. The mean track lengths of RN (38.29 steps) and HN (30.64 steps) clearly differ from each other, even though the test sounds in both were the same.

Fig. 29 shows the distribution of track lengths per condition and per subject. The shortest track in the whole experiment was only 19 steps (4.8 minutes) long, whereas the longest tracks were 50 (11.2 minutes) steps, defined by the hard-coded limit in the experimental setup. The average track length was 34.65 steps (8.2 minutes). By visual inspection both figures seem to follow normal distribution. Overall, the tracks lengths were well distributed across subjects and conditions.

The track length distributions were analyzed with the repeated-measures ANOVA. The distributions were spherical (Mauchly's Test of Sphericity: $\chi^2_{(27)} = 28.946$, p = 0.400), thus no sphericity correction was applied. The ANOVA returned non-significant results both for within subject effects ($F(7,1) = 1.647$, $p = 0.132$) and between subject effects ($F(7,1) = 1636.52$, $p = 0.000$). In other words, the distributions did not differ.

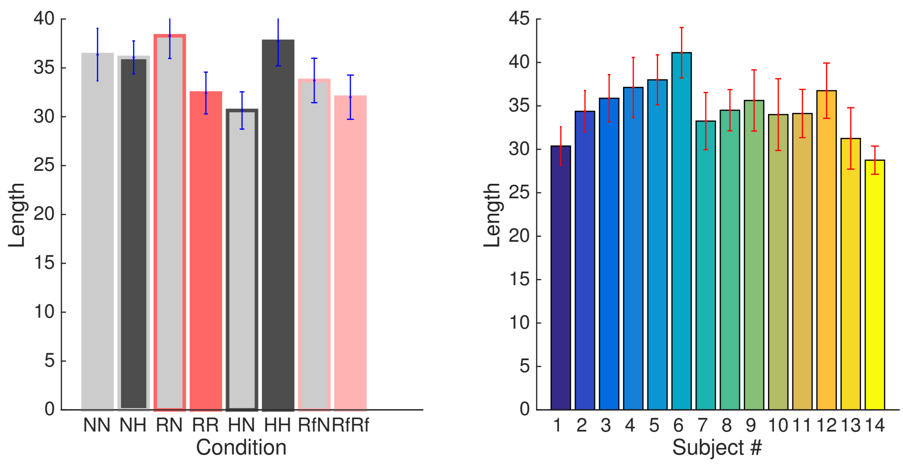

Figure 28: Left: track length mean values per condition over all 14 subjects. Right: track length mean values per subject over all eight conditions. The error bar denotes one standard error.

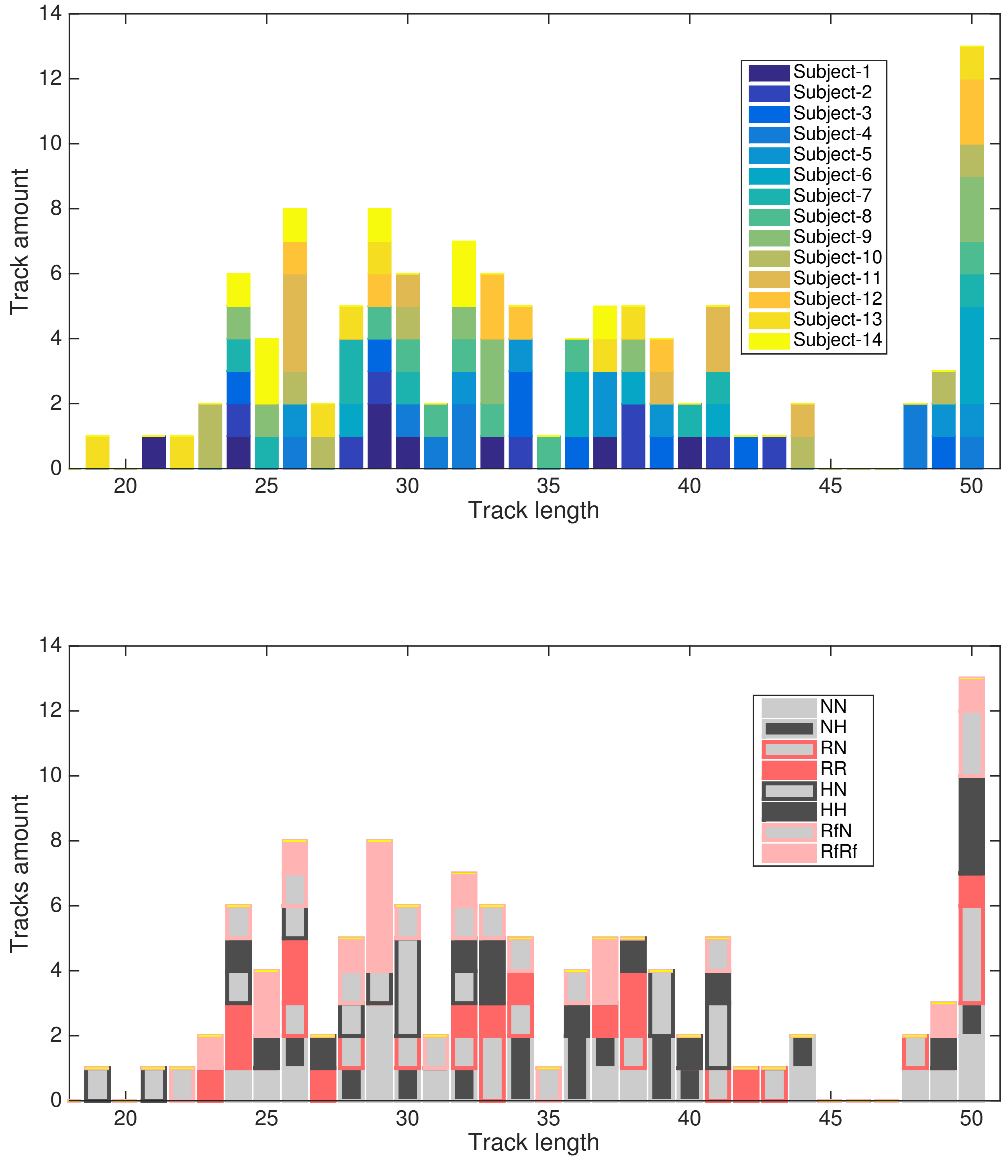

Figure 29: Distribution of the track lengths in the whole experiment per condition and per subject.

4.4.3 Response time course

Fig. 30 shows the time course of the response mean values per condition and per subject. The figure also includes the theoretical time courses showing the measured average evolution, the best possible track produced if all answers were correct and the simulated track produced if all responses were pure guesses. The simulated track, produced in Python programming language by averaging 1000 simulation of the experiment, is based on the assumption that the probability of giving the correct response by guessing is $\frac{1}{3}$. This is because there were three response options from which one was correct. Thus, the probability of getting all the 19 steps correct by guessing is extremely small, $(\frac{1}{3})^{19}$. By comparing the measured mean track with the simulated one it becomes clear that the responses given by the subject were not pure guesses. By purely guessing 19 first steps would shift the threshold level up, close to the value of 30 dB. During the first steps the mean track is very close to the best possible track, in other words, the subjects can easily hear the test sound. This means that the initial threshold chosen for the method could be easily 3–6 dB lower than the currently chosen 20 dB.

Inspecting the time course of the condition means in Fig. 30 reveals that all tracks converge to a narrow range of about 11.5–13 dB during the first 19 steps. Fig. 31 shows the zoomed-in version of the condition tracks around 2 minutes. RR, RfN and RN seem to stand out when they are compared with the mean track: RR shows clearly lowered thresholds whereas RfN and RN show elevated thresholds. The evolution of tracks per subject in Fig. 30 shows clear subject-wise differences in convergence. The highest value at step 19 was 17.13 dB by subject #3 and the lowest value 8.63 dB by subject #3. The mean value was 12.19 dB. Tracks of subjects #3 and #9 follow the best possible tracks consistently in every condition for the 7 first steps.

Condition mean values were calculated both for the final thresholds and the different step ranges of the adaptive tracks. These are shown in Fig. 32. Conditions in which the inducer and the test sound were the same seem to have lower thresholds compared to the most of the respective cases where the test sound was different than the inducer (NN-vs-NH, RN-vs-RR, HN-vs-HH, RfN-vs-RfRf). However, the error margin in the means is considerably high, thus no conclusions can be drawn based on them.

Overall, the adaptive tracks converge to a certain final threshold non-monotonically showing exponential decay. In the beginning of the tracks the subjects tend to give wrong answers, probably because they did not pay attention to the interface. However, after this the threshold typically starts going down. The subjects tend to reach exceptionally low thresholds when going down for the first time. But as soon as they answer wrong the measured threshold bounces upwards for several consecutive trials, after which it turns down again. However, this is considered typical for threshold measurements [13].

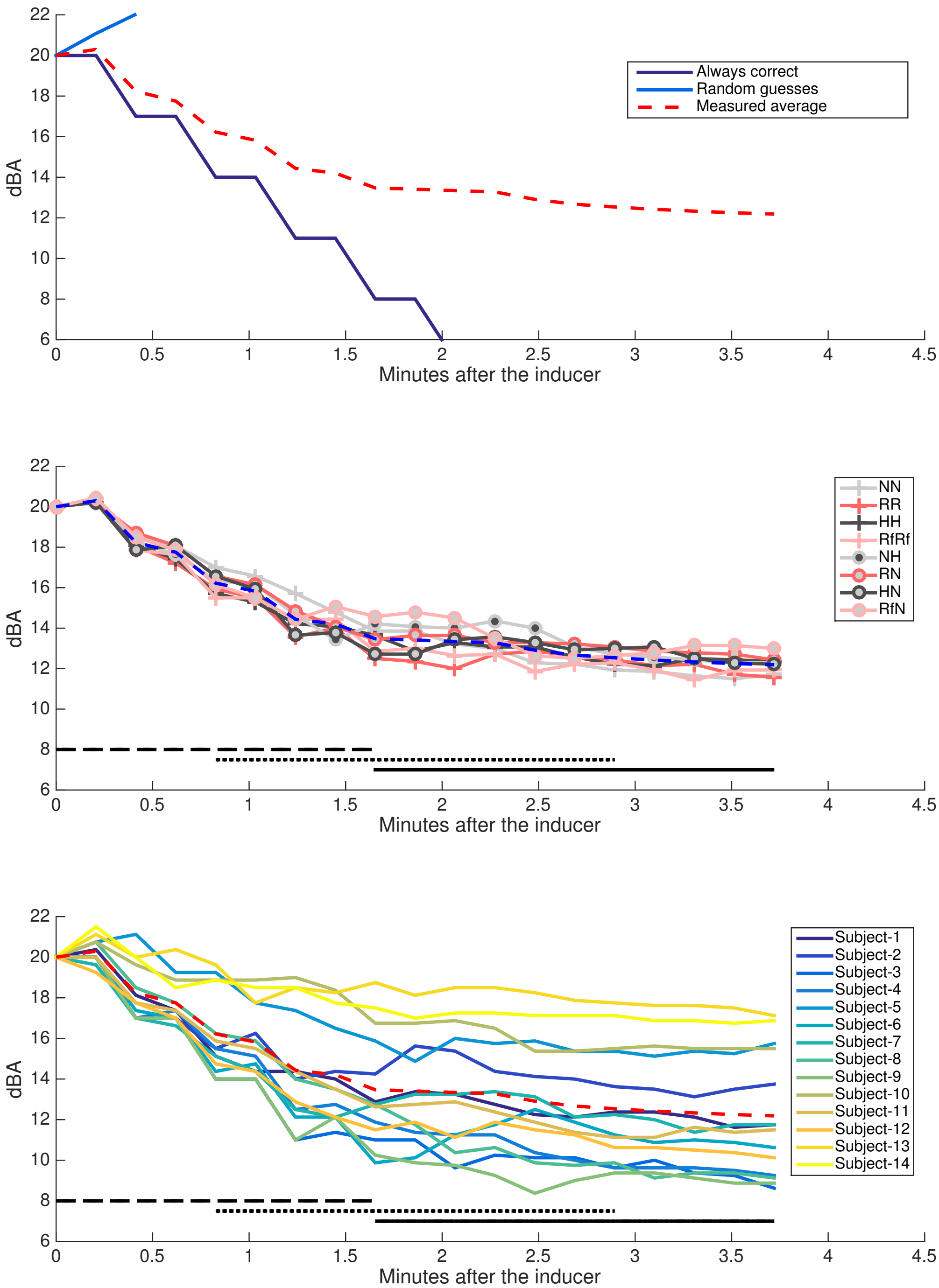

Figure 30: Time courses of the adaptive tracks. Top: the track produced by random guesses, the measured mean track and the theoretical track if all responses were correct. Middle: condition means over subjects. Bottom: subject means over conditions. The black lines show different step ranges or the data sets: steps 1-10 (dashed line), steps 5-15 (dotted line), steps 9-19 (solid line). The red and blue dashed lines represent the mean of all tracks.

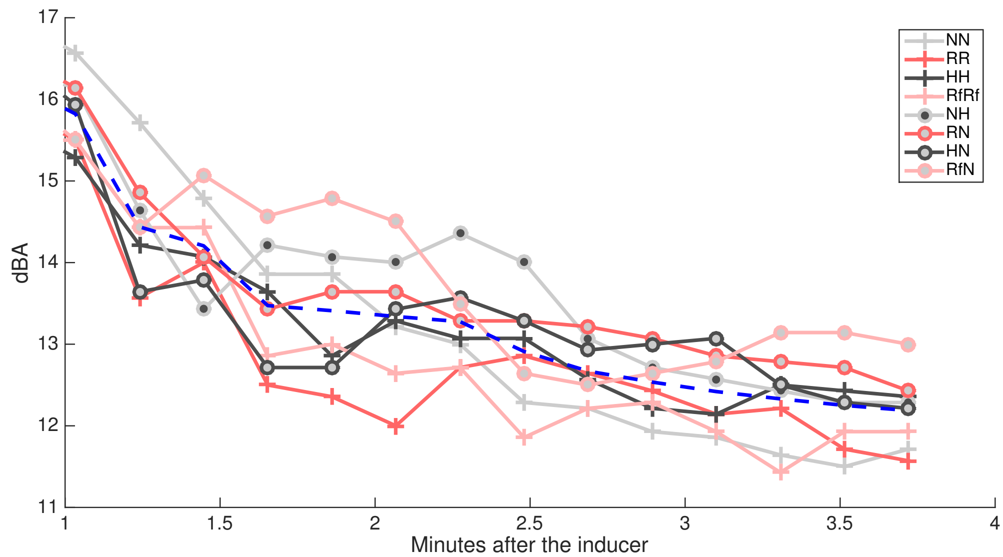

Figure 31: A closer look at the region where the condition means differed the most. The blue dashed line represents the mean of all tracks.

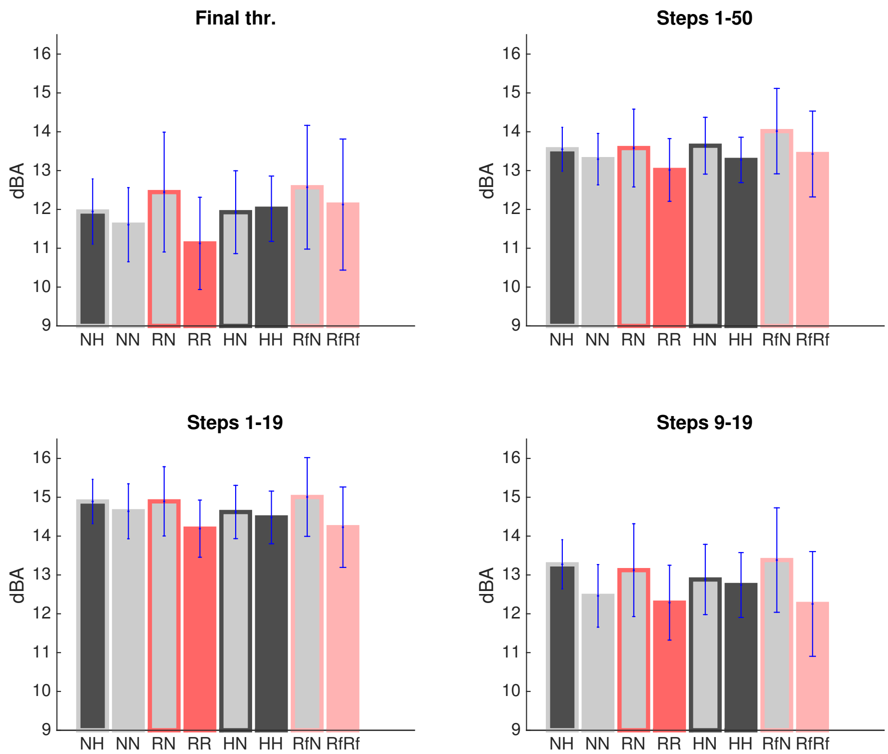

Figure 32: Mean thresholds of the conditions in four different analysis ranges. The error bars denote the standard errors.

4.4.4 Final thresholds and audiometry

The mean final thresholds were compared subject-wise with the pure-tone averages (PTA) calculated from the audiometry results. This was done to see whether there is correlation, i.e. if the PTAs predict the mean final thresholds. The PTAs plotted against the mean final thresholds are shown in Fig. 33. Both the PTAs and the mean final thresholds were tested for normality using Shapiro-Wilk and Kolmogorov-Smirnov normality tests. Since both tests returned a p-value greater than 0.05, normality cannot be excluded, and thus the correlation was tested with the Pearson test instead of the Spearman test. The test found no correlation between the PTAs and the mean final thresholds (Pearson: $r^2_{(14)} = 0.123$, p = 0.676).

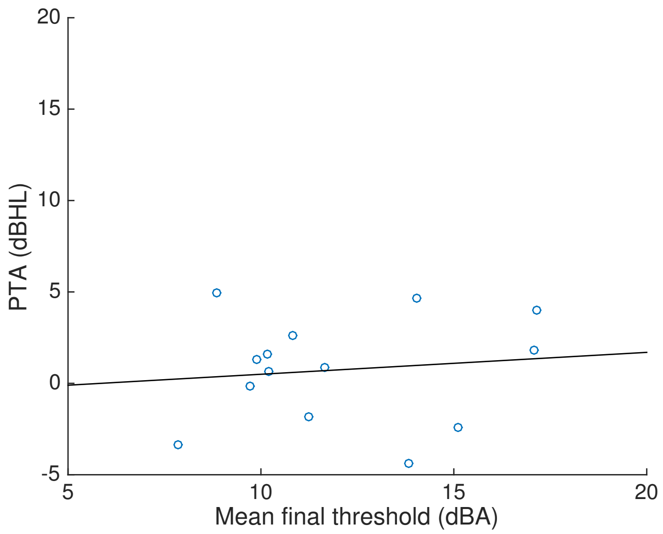

Figure 33: Averaged audiometry result over frequencies and ears versus averaged final threshold result over all eight conditions. Each point represents a subject. There is no significant correlation between the results, indicated by the line (Pearson: $r^2_{(14)} = 0.123$, p = 0.676).

The Pearson correlation test was also run separately for the final thresholds of each condition and the audiometry results of different frequencies. The correlation coefficients and p-values are shown in Table 5. No significant correlations were found when all 14 subjects were included. However, when subjects #5 and #10 were excluded as outliers, significant positive correlations were found between the 4 kHz audiometry results and the final thresholds of conditions NN, HN and HH.

Table 5: Pearson correlation of condition-specific final thresholds and different audiometry results: pure tone averages (PTA), 1 kHz, 2 kHz and 4 kHz. All the results were averaged over both ears. All 14 subjects were included in the calculation of r^2_{14}. Subjects #5 and #10 are excluded in the calculation of r^2_{12}. The significance level $p \leq 0.05$ is used.

	PTA		1 kHz		2 kHz		4 kHz			
Cond.	r^2_{14}	p	r^2_{14}	p	r^2_{14}	p	r^2_{14}	p	r^2_{12}	p
NN	-0.015	0.958	-0.0577	0.845	0.0406	0.891	0.498	0.070	0.695	**0.012**
NH	-0.191	0.513	0.188	0.521	-0.215	0.461	0.029	0.921	0.547	0.655
RN	0.189	0.518	0.0212	0.953	0.163	0.577	0.451	0.105	0.504	0.095
RR	0.240	0.410	0.332	0.246	0.0143	0.961	0.322	0.262	0.472	0.121
HN	-0.010	0.974	0.0227	0.939	0.158	0.589	0.325	0.257	0.623	**0.030**
HH	-0.082	0.781	0.167	0.567	-0.0149	0.960	0.078	0.791	0.608	**0.036**
RfN	0.187	0.523	0.179	0.542	0.254	0.381	0.322	0.261	0.443	0.149
RfRf	0.245	0.399	0.173	0.554	0.359	0.208	0.390	0.162	0.549	0.065

4.4.5 Statistical analysis

All six data sets listed earlier in Table 2 were analyzed using the One-Way repeated-measures ANOVA, introduced in section 2.2.5. Final thresholds contain the average dB-values of the last four turning points of the adaptive tracks. The rest of the five data sets contain the dB-values of the adaptive tracks on a trial-by-trial basis. The missing trials in the data set "steps 1-50" wdere treated as missing answers.

The sphericity of the data sets was tested with the Mauchly's Test of Sphericity, the results of which are shown in Table 6. According to the test none of the sets are spherical which is why Greenhouse-Geisser corrections were applied to the ANOVA. Table 6 also shows the results of the within and between subject tests of the One-Way repeated measures ANOVA. Within subjects factors are the subjects themselves and the steps of the tracks. Between subject factors are the conditions.

Post-hoc tests with Bonferroni correction applied were done in order to find differences between the conditions. The tests revealed significant differences in 7 out of the 28 comparison pairs done. The significant differences, defined as the significance level $p \leq 0.05$ after correction, are shown in Tables 7, 8 and 9. All non-significant cases, except RR-NN and HH-HN, are intentionally left out.

Note that there were no significant differences found between any conditions when the final thresholds were analyzed, and therefore they are not shown in the tables. The analysis of different steps ranges also revealed that no differences were found in steps 1-10 and 5-15 alone. Instead all differences were in steps 1-50 (6 differing conditions), 1-19 (3 differing conditions) and 9-19 (2 differing conditions).

Table 7 shows the significant differences between the conditions where one contained only aftereffect-eliciting sounds (RR) and the other only sounds known not to cause the aftereffect (NH, HN, HH). All differences were found on a wide time scale,

Table 6: Since all p-values from Mauchly's Test of Sphericity are smaller than 0.05, none of the data sets are spherical. This is why Greenhouse-Geisser corrections were applied to the within and between subject tests.

	Mauchly's Test of Sphericity		Within subjects tests		Between subjects tests	
	$\chi^2_{(27)}$	p	F	p	F	p
final thr.	386.30	0.000	F(2.78,1)=0.68	0.561	F(2.78,1)=226.57	0.000
steps 1-50	443.12	0.000	F(5.08,1)=4.25	0.001	F(5.08,1)=5038.87	0.000
steps 1-19	59.21	0.001	F(4.84,1)=3.25	0.007	F(4.84,1)=4590.89	0.000
steps 1-10	192.80	0.000	F(4.87,1)=2.35	0.041	F(4.87,1)=3871.80	0.000
steps 5-15	224.86	0.000	F(4.81,1)=2.11	0.065	F(4.81,1)=3720.18	0.000
steps 9-19	296.52	0.000	F(4.45,1)=3.21	0.010	F(4.45,1)=3173.40	0.000

in the step range 1-50. However, against expectations, conditions RR and NN did not differ.

Table 8 lists the significant differences between conditions in which the inducer sounds were the same but the test sound varied. By looking at the table, it looks like using the same inducer and the test sound (NN, RR, RfRf) yields lower thresholds than using different sounds (NH, RN, RfN), with one exception: if the HH followed the general trend, its DT should be significantly lower than the DT of HN. Another irregularity can be seen in the time course. The difference in NN-vs-NH can only be seen during steps 9-19, corresponding to about 2-4 minutes after the inducer sound, whereas in other comparisons (RR-vs-RN, RfRf-vs-RfN) the differences were detected on a wider time scale. This can indicate that the effect was stronger when R or Rf was used as the inducer than when N was used.

Table 9 shows one more significant difference (RR-vs-RfN) found on a wide time scale, in steps 1-50. In further discussions this result will be discarded since it was not consistent with other differences found.

Table 7: The significant differences ($p \leq 0.05$, bold and underlined) found between the conditions where the other contained an aftereffect-eliciting inducer and the other did not. There were no significant differences when only the final thresholds were analysed.

	Analyzed steps				
	1-50	1-19	1-10	5-15	9-19
RR-vs-NH	**<u>0.005</u>**	0.074	1.000	0.187	0.092
RR-vs-HN	**<u>0.030</u>**	1.000	1.000	1.000	1.000
RR-vs-HH	**<u>0.002</u>**	1.000	1.000	1.000	1.000
RR-vs-NN	0.895	1.000	0.347	1.000	1.000

Table 8: The significant differences ($p \leq 0.05$, bold and underlined) found between the conditions in which the inducer sounds were the same. There were no significant differences in final thresholds.

	Analyzed steps				
	1-50	1-19	1-10	5-15	9-19
NN-vs-NH	1.000	1.000	1.000	1.000	**<u>0.011</u>**
RR-vs-RN	**<u>0.040</u>**	**<u>0.047</u>**	0.789	0.483	0.331
RfRf-vs-RfN	**<u>0.001</u>**	**<u>0.002</u>**	0.636	0.163	**<u>0.001</u>**
HH-vs-HN	1.000	1.000	1.000	1.000	1.000

Table 9: Other significant differences ($p \leq 0.05$, bold and underlined) found. There were no significant differences when only the final thresholds were analysed.

	Analyzed steps				
	1-50	1-19	1-10	5-15	9-19
RR-vs-RfN	**<u>0.020</u>**	**<u>0.043</u>**	0.343	0.579	0.130

4.5 Discussion on results

The results of the audiometry and the actual experiment were reported previously in sections 4.3 and 4.4. In this section those results and their potential causes are discussed further. The research questions and the hypothesis stated in section 3.1 are also evaluated.

Button presses. Inspecting the button hit rates in the AFC part of the experiment revealed the rate of button 2 to be about 10% higher than the rate of buttons 1 and 3. In other words, the subjects managed to answer button 2 correctly more often than the two other buttons. This can indicate several things. One interpretation is that the sound in button 2 was easier to detect, in other words, the hearing would be more sensitive 500-700 ms after the inducer. However, there are no previous studies supporting this. The most probable explanation lies in the answer behavior of the subjects: the subjects most probably avoided pressing button 2, located in the middle of the interface, in unsure situations.

Track lengths. When the track length distributions were analyzed with the repeated-measures ANOVA, no significant differences were found. However, the track length means per condition (Fig. 28) show interesting differences between conditions with the same inducer sound. The track lengths in NN-vs-NH and RfN-vs-RfRf are similar, implying that the test sound was equally easy to find. However, the other two condition pairs differ: the track length of RN is clearly longer than RR, and HN clearly shorter than HH.

The mean track lengths of RN (38.29 steps) and HN (30.64 steps) clearly differ from each other, even though the test sounds in both conditions were the same. This strongly suggests that exposure to R caused the DT of N to fluctuate more than exposure to H. However, since the DT of N was not tested without the inducer, the absolute level of fluctuation is not known. Thus, it cannot be concluded whether R increased the amount of fluctuation or whether H reduced it.

Final thresholds and audiometry. A significant positive correlation was found between the 4 kHz audiometry results and the final thresholds of almost all aftereffect-free conditions (NN, HN, HH). The final threshold of one aftereffect-free condition (NH) did not correlate with the audiometry result, which may indicate that N altered the detection of H. In other words, the shape of the hearing sensitivity curve measured in silence changed in NH, whereas in NN, HN and HH it remained unchanged.

The fact that the correlation was found at 4 kHz may indicate that the detection of N and H at the hearing threshold rely on frequencies at around 4 kHz. This would be in line with what is known about hearing, introduced earlier in section 2.1: under normal circumstances the hearing is the most sensitive at the frequency range 2–4 kHz [4, Chapter 6]. However, there are also arguments against this explanation. First, against expectations no significant correlations were found at the adjacent frequency, 2 kHz, and second, the headphones used in the DT measurement affected the DTs because their frequency response was not flat.

The correlation at 4 kHz seems to be tied to the absence of the aftereffect: when the final threshold of N was measured under the influence of an aftereffect (RN, RfN), there was no significant correlation at any frequency. This is an indication of the aftereffect manipulating the DT at 4 kHz so that it becomes different from the DT measured in silence with audiometer. However, this does not necessarily mean that the DT of wide-band stimuli would be altered.

Several audiometry frequencies were tested in order to find correlations with final thresholds but only the abovementioned correlations were significant. A separate correlation test was also ran for those subjects who did the audiometry just before the actual experiment. Since these subjects underwent the audiometric screening on a different day than the actual experiment, and it is known that hearing sensitivity can change during the day, e.g. after exposure to loud noises [4, Chapter 31], some differences in the test results were expected. However, no correlation was found. One possible explanation for this is the fact that two different types of headphones with different frequency responses were used in the audiometric screening and in the DT measurement. In addition, the hearing sensitivity in silence and after exposure to noise-like stimuli could differ per se. Overall, comparing the audiometry results with the experiment results shows how little pure-tone audiometry results tell about how we hear everyday sounds in everyday conditions.

Is the DT different after exposure to an aftereffect-eliciting sound compared to an aftereffect-free sound? The DT was significantly lower only in one aftereffect-eliciting condition (RR) when compared to aftereffect-free conditions (NH, HH, HN). These differences occurred at steps 1–50, i.e. 0–10.1 minutes after the inducer. However, as an exception, no difference was found between RR and the only remaining aftereffect-free condition, NN. One could come to the conclusion that the RR-vs-NH and RR-vs-HN differences are explained by the use of the same sound as the inducer and as the test sound. But this cannot be the only explanation since RR-vs-HH shows no difference. Against expectations, there was no contrast between the other aftereffect-eliciting condition, RfRf, and any of the aftereffect-free conditions. This and the fact that the difference was distributed over the whole measurement range, steps 1–50, would indicate that the difference seen with RR is more likely related to the R sound itself than directly to the aftereffect.

As stated earlier in the hypothesis, conditions not involving the aftereffect (NN, NH, HH, HN) were expected to have similar DTs. In fact, the statistical tests show this to be true with the exception of NH being significantly higher than NN at steps 9–19 (Table 8), i.e. 1.7–3.7 minutes after the inducer. This is also visible in Fig. 31. There is yet another factor, mentioned earlier, that differentiates NH from the rest: the final threshold of NN, HN and HH show significant positive correlation with the 4 kHz audiometry results, whereas NH does not.

Does listening to a specific sound alter the DT of that sound only? Based on the results of the statistical tests (see Table 8), it generally looks like using the same test sound as the inducer (NN, RR, RfRf) tends to yield lower thresholds than using a different test sound (NH, RN, RfN). Visual inspection of the condition means in Fig. 32 reveals that the differences in condition means occur already within 4 minutes after the exposure and are typically subtle, less than 1 dB between the conditions. HN-vs-HH forms an exception: the DT in HH does not differ significantly from that in HN. As mentioned before, the adaptive track of HH, seen in Fig. 29, is also on average longer than the track of HN, indicating that there was more sensory noise present in HH, and thus the detection of H was harder. Other conditions with the same inducer and test sound produced shorter or equally long tracks as the corresponding condition containing a different test sound.

The exception, HN-vs-HH, raises a question of how widely the general trend can be generalized. Future studies are needed in order to find out why the thresholds were lowered for NN, RR and RfRf and why it did not occur with HH. For example, measuring the DTs of N and H without the inducer would help in drawing more conclusions. There is also an irregularity occurring in the time course of the conditions that follow the trend (NN, RR, RfRf): the NN-vs-NH differences can only be seen during steps 9–19, corresponding to about 2–4 minutes after the inducer sound, whereas in other conditions (RR vs RN, RfRf vs RfN) the difference was spread over the whole range of steps 1–50.

How does the time course of the DTs behave after exposure to the stimuli? The time course of the first steps, up to about half a minute, are very similar across conditions, following almost the best possible "always correct" track (Fig. 30). After the first steps the condition mean tracks start to diverge and then converge towards their final thresholds. As a general trend the convergence of the conditions means happens almost monotonically, showing exponential decay over the 50 steps in the whole experiment, of which NN is a good example. The zigzag shape of the track, e.g. in RfRf, is mainly because of the method. Namely, it is known that the subjects tend to reach low thresholds when going down in the threshold measurement, especially for the first time. But as soon as they answer wrong the measured threshold bounces upwards for several consecutive trials, after which it turns down again, and continues to zigzag in this way towards the final threshold. When the final threshold of the conditions were statistically tested, no differences were found between them. There are at least two plausible explanations for this: either the termination rule, "six consecutive turning points within a 4-dB range", did not stop in the right place and is in need of adjustment, or the refresher sounds were not effective enough in keeping the effect on.

Compared to the normal type of time course showing nearly monotonic exponential decay, RfN and NH stand out at around 2 minutes, see Fig. 31. Their thresholds bounce upwards for several consecutive steps after the first descent. The maximums of the bounces are almost 2 dB higher than the final converged thresholds. In section 3.1 this kind of bounce was hypothesized only for R and Rf, however, according to the results it is also present in NH.

5 General discussion

This research had three goals. The first was to find out if the detection threshold (DT) is different after exposure to a Rosenblith aftereffect-eliciting sound in comparison to an aftereffect-free sound. The second goal was to examine if listening to a specific sound alters the DT of that sound only. The third goal was to analyze the time course of the DTs after exposure to the stimuli.

Is the DT different after exposure to an aftereffect-eliciting sound compared to an aftereffect-free sound? In the present study the aftereffect was not clearly related to altered DTs. The DT was significantly different (lower), 0–10 minutes after the inducer, only in one aftereffect-eliciting condition when compared to aftereffect-free conditions. However, this finding cannot be generalized since another condition known to induce the aftereffect was not different from any other condition. To draw further conclusions about the relationship between the DTs and the Rosenblith aftereffect, the DTs of the test sounds without a preceding exposure should be measured.

Some indications of aftereffect-related alteration in the DT of pure tones were found. Namely, when no aftereffect was induced prior to testing, the final DT of aftereffect-free noise correlated with the DT of the 4 kHz pure tone measured with audiometer. However, in conditions in which the aftereffect was induced prior to testing, no correlation was found, indicating a possibility that the aftereffect altered the DT of 4 kHz from that measured in silence. All in all, it is still rather unclear whether the Rosenblith effect induced by the pulse train is related to altered DTs of noise-like stimuli.

Does listening to a specific sound alter the DT of that sound only? Exposure to a steady sound was found out to significantly lower the DT of that specific exposure sound compared to some other exposure sounds. However, the difference was typically less than 1 dB, and, based on a previous study [46] on DTs, probably not even a sign of hypersensitivity. Despite of this, the lowered threshold, that was observed, is contradictory to the elevated thresholds hypothesized by the theory of adaptation. If auditory adaptation were the prevailing process in the hearing system after the exposure, opposite differences would have been expected. One possible explanation for the differences is that listening to the inducer caused the test sound to become familiar to the subject. In other words, the test sound was detected at a lower SPL because the subjects knew what to look for. This explanation could be tested by letting the subjects familiarize themselves with the sounds in preceding training sessions. As the training proceeds, the differences are expected to slowly fade away, as happened e.g. for the detection of AM in [32].

The time course of the lowered thresholds was not the same for all stimuli. The lowered threshold for noise was present about 2–4 minutes after the exposure, whereas the lowered threshold after the aftereffect-eliciting stimuli were generally measured throughout the range 0–10 minutes. Whether the cause of the difference was related more to the aftereffect or to the effect of using the same inducer and test sounds could be examined by measuring the DTs of the stimuli without preceding exposure.

How does the time course of the DTs behave after exposure to the

stimuli? As expected with the use of the adaptive tracking method, the measured DTs showed exponential decay over several minutes on the way towards their final thresholds. No differences were found between the final thresholds of conditions, indicating that any adaptation or fatigue elicited by the exposure faded away in less than 10 minutes. An alternative explanation is that the effect was similar across conditions, thus no differences were observed. There were also no differences right after the exposure. However, this is due to the method, ultimately by the use of a too high initial SPL (20 dB) for the test sounds. But only because the method could not grasp any differences in DTs between the conditions during the first steps does not mean that there would not be any. To grasp those differences the method should be adjusted to start at a lower test tone SPL.

In two conditions, including both an aftereffect-eliciting and an aftereffect-free condition, an unexpected bounce-like elevation in the noise detection threshold was measured around 2 minutes after the exposure. The bounces, whose maximum value is 2 dB higher than the value of the final converged thresholds, indicate a possible connection with the two-minute bounce found in previous studies [33, 47, 34]. In a study [34], that used only pure tones as inducers, loud low frequency tones were found to induce the bounce the best. It was argued that in order for the bounce to occur the stimulus had to excite a large expanse of the cochlea which is what the loud low-frequency tones are capable of. The study did not report whether the bounce was more because of the spreading excitation pattern or the high SPL. Either way, the broad-band inducers in the experiment of this thesis also have the ability to excite the cochlea widely. Thus, it is highly likely that the increase in threshold seen both after an aftereffect-eliciting exposure and an aftereffect-free exposure is the same as the two-minute bounce in [33]. However, why there was no increase with the other inducers remains an open question, to which the answer might lie in the imperfections of the present experimental method.

5.0.1 Improvement ideas

Several ways of improving the method of the experiment have been identified in retrospect. The first improvement concerns the audiometry results that were found to be biased (section 4.3, Fig. 27). Suggested methods to prevent the bias are familiarization trials and counterbalancing.

Second, the detection of modulation in a short test tone of 100–300 ms is questionable. The average test sound duration was 200 ms which is enough, e.g. for playing back 20 pulses at the rate of 100 Hz (R). However, if the test sound is modulated with a relatively low frequency, e.g. 4 Hz (Rf), the average duration of the test sound is too short even for one modulation cycle to occur. One solution, although not trade-off-free, would be to use a higher modulation frequency or to make the test sound longer.

The third improvement is related to the termination rule of the adaptive track. In the experiment no significant differences were found between the converged final thresholds, which was probably due to the aftereffect fading away over time. To keep the effect on, the refresher sound should be made longer or intenser.

As the fourth improvement the test sounds should also be tested without inducers. The knowledge on the DTs of the stimuli alone, without the preceding inducer, would be useful in differentiating between the inducer-related effects and the effects related to the test sound itself.

Lastly, since the greatest differences across conditions seem to lie in the region of about two minutes after the inducer (Fig. 31), the method should be adjusted to better grasp that time range. This could be achieved with the following improvements:

1. Pre-test the stimulus DTs for each subject and stimulus. Start the conditions closer to the individual thresholds, in the same way as in preliminary tests in the method of constant stimuli. As mentioned previously, the initial threshold could be easily 3–6 dB lower than the currently chosen 20 dB. The first steps right after the exposure can be crucial for finding differences between conditions, thus, it is of utmost importance of not limiting them by the method.
2. Use 2 dB step size instead of 1 dB, as in [28]. This ensures that the appropriate range is achieved faster.
3. Use one-up one-down adaptive procedure with 9-AFC. In the current one-up two-down 3-AFC procedure going one step down takes 24.8 seconds and there is a $\frac{1}{9}$ chance of achieving it by guessing. Using one-up one-down with 9-AFC, going one step down would only take 14.8 seconds and the chance of guessing would remains the same, $\frac{1}{9}$. However, using 9-AFC could cause other kind of problems, e.g. subjects forgetting the number of the flashing button during which he or she heard the sound.

5.0.2 Future research

DT of pure tones after exposure. By comparing correlations of the audiometry results and the DTs of the stimuli, evidence was found to claim that the aftereffect alters the detection of a 4 kHz pure tone from that measured in silence. The detection of pure tones under the influence of the aftereffect in general could be worth further studies.

The effect of modulation. Although the stimuli tested in the present experiment included a modulated sound (Rf), testing the effect of modulation on DTs was not the main goal of the study. As a matter of fact, the effect of modulation on the DTs of sounds to which they were applied is an understudied topic. The studies related to modulation and DTs are mostly restricted to temporal modulation transfer functions, i.e. how the DT of modulation itself depends on modulation frequency and modulation depth. One could hypothesize, based on the theory that modulation in the sound signal activates separate channels in the auditory system [31], that any signal on which modulation has been applied to would yield into lower DTs also for the sound itself, compared to the unmodulated version of the same sound. Studies on modulation are closely related to the Rosenblith aftereffect itself, since the aftereffect is suggested to elicit changes in the detection of AM [28].

Perceptual aspects. The stimuli (H, R, Rf) in the present study included both high modulation frequencies (R, Rf: 100 Hz), perceived as rough, regardless of their periodicity [12, Chapter 11], and low modulation frequencies (Rf: 4 Hz) that elicit the sensation of fluctuation. There are many open questions related to the timbre of the aftereffect-eliciting sounds. For example, the dependency between the fluctuation strength or roughness and the Rosenblith aftereffect is unknown – so is the exact description of the aftereffect missing.

Another perceptual aspect closely related to the sensation of fluctuation and roughness is the sensation of annoyance, and its relationship to the DTs of the stimuli. Based on both the statistical tests and the inspection of the condition mean values, the informal subject-reported annoyances of the stimuli and their DTs do not seem to have a clear relationship. In other words, the results of this study do not directly support the claim that more annoying sounds would be detected at a lower SPL. However, future studies on this are needed as well.

6 Conclusions

In the present study the aftereffect was not clearly related to altered noise detection thresholds of the tested wide-band stimuli. However, some indications of a change in the hearing threshold curve after the exposure to an aftereffect-eliciting sound were found. Namely, when no aftereffect was induced prior to testing, the detection thresholds of aftereffect-free noise correlated with the detection threshold of the 4 kHz pure tone measured by audiometry. However, when the aftereffect was present, no correlation was found, implying that the aftereffect alters the noise detection threshold from that measured in silence.

The sound that the subject was just exposed to, was generally detected at a lower threshold, in some cases up to 1 dB lower, than some other sound. This is contradictory to the elevated threshold hypothesized by adaptation, and indicate that higher level processes take place in the detection.

The recovery of the noise detection threshold after the exposure showed exponential decay. However, an unexpected bounce-like elevation in the detection threshold was present in two of the tested eight conditions, including both an aftereffect-eliciting and an aftereffect-free condition, at around 2 minutes after the exposure. These bounces, whose maximum value is 2 dB higher than the value of the final converged thresholds, are most probably the same "two-minute bounce" phenomenon found in previous studies [33, 34, 47].

References

[1] R Addams. LI. An account of a peculiar optical phænomenon seen after having looked at a moving body. *The London and Edinburgh Philosophical Magazine and Journal of Science*, 5(29):373–374, 1834.

[2] E Zwicker. "Negative afterimage" in hearing. *The Journal of the Acoustical Society of America*, 36(12):2413–2415, 1964.

[3] WA Rosenblith, GA Miller, JP Egan, IJ Hirsch, and GJ Thomas. An auditory afterimage. *Science*, 106:333–335, 1947.

[4] TD Rossing. *The Science of Sound*. Addison Wesley Publishing Company, 1990.

[5] VR Algazi, RO Duda, DM Thompson, and C Avendano. The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 99–102. IEEE, 2001.

[6] V Pulkki and M Karjalainen. *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*. John Wiley & Sons, 2015.

[7] F Dunn, WM Hartmann, DM Campbell, NH Fletcher, and Thomas Rossing. *Springer handbook of acoustics*. Springer, 2015.

[8] M Karjalainen. *Kommunikaatioakustiikka*. Helsinki University of Technology, 1999.

[9] S Bech and N Zacharov. *Perceptual audio evaluation - Theory, method and application*. John Wiley & Sons, 2007.

[10] HCCH Levitt. Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B):467–477, 1971.

[11] ISO Norma. 8253-1: 2010 Acoustics-Audiometrics test methods.

[12] E Zwicker and H Fastl. *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media, 2013.

[13] WH Ehrenstein and A Ehrenstein. Psychophysical methods. In *Modern techniques in neuroscience research*, pages 1211–1241. Springer, 1999.

[14] JF Hair. *Multivariate Data Analysis: A Global Perspective. 7th ed.* Prentice Hall, 2009.

[15] A Field. *Discovering statistics using SPSS*. Sage, 2005.

[16] GV Glass, PD Peckham, and JR Sanders. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42(3):237–288, 1972.

[17] SM Scariano and JM Davenport. The effects of violations of independence assumptions in the one-way ANOVA. *The American Statistician*, 41(2):123–129, 1987.

[18] O Räsänen and UK Laine. Time-frequency integration characteristics of hearing are optimized for perception of speech-like acoustic patterns. *The Journal of the Acoustical Society of America*, 134(1):407–419, 2013.

[19] SP Bacon and NF Viemeister. Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. *Audiology*, 24(2):117–134, 1985.

[20] WD Ward. Effects of noise on hearing thresholds. *Noise and Public Health Hazard. OSHA Report*, 4, 1969.

[21] IJ Hirsh and RC Bilger. Auditory-threshold recovery after exposures to pure tones. *The Journal of the Acoustical Society of America*, 27(6):1186–1194, 1955.

[22] D McFadden and HS Plattsmier. Frequency patterns of TTS for different exposure intensities. *The Journal of the Acoustical Society of America*, 74(4):1178–1184, 1983.

[23] RH Wilson, JF Steckler, HC Jones, and RH Margolis. Adaptation of the acoustic reflex. *The Journal of the Acoustical Society of America*, 64(3):782–791, 1978.

[24] RH Margolis and GR Popelka. Loudness and the acoustic reflex. *The Journal of the Acoustical Society of America*, 58(6):1330–1332, 1975.

[25] BCJ Moore. *An introduction to the psychology of hearing: BRILL.* 2012.

[26] SP Bacon and DW Grantham. Modulation masking: Effects of modulation frequency, depth, and phase. *The Journal of the Acoustical Society of America*, 85(6):2575–2580, 1989.

[27] JD Harris and AI Rawnsley. The locus of short duration auditory fatigue or "adaptation". *Journal of Experimental Psychology*, 46(6):457, 1953.

[28] A Gutschalk, C Micheyl, and AJ Oxenham. The pulse-train auditory aftereffect and the perception of rapid amplitude modulations. *The Journal of the Acoustical Society of America*, 123(2):935–945, 2008.

[29] M Sams, R Hari, J Rif, and J Knuutila. The human auditory sensory memory trace persists about 10 sec: neuromagnetic evidence. *Journal of Cognitive Neuroscience*, 5(3):363–370, 1993.

[30] N Ulanovsky, L Las, D Farkas, and I Nelken. Multiple time scales of adaptation in auditory cortex neurons. *The Journal of Neuroscience*, 24(46):10440–10453, 2004.

[31] BW Tansley and JB Suffield. Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation. *The Journal of the Acoustical Society of America*, 74(3):765–775, 1983.

[32] L Bruckert, M Herrmann, and C Lorenzi. No adaptation in the amplitude modulation domain in trained listeners. *The Journal of the Acoustical Society of America*, 119(6):3542–3545, 2006.

[33] IJ Hirsh and WD Ward. Recovery of the auditory threshold after strong acoustic stimulation. *The Journal of the Acoustical Society of America*, 24(2):131–141, 1952.

[34] DL Kirk and RB Patuzzi. Transient changes in cochlear potentials and DPOAEs after low-frequency tones: the 'two-minute bounce' revisited. *Hearing Research*, 112(1):49–68, 1997.

[35] JR Hughes. Auditory sensitization. *The Journal of the Acoustical Society of America*, 26(6):1064–1070, 1954.

[36] JR Hughes and WA Rosenblith. Electrophysiological evidence for auditory sensitization. *The Journal of the Acoustical Society of America*, 29(2):275–280, 1957.

[37] JP Franosch, R Kempter, H Fastl, and JL van Hemmen. Zwicker tone illusion and noise reduction in the auditory system. *Physical Review Letters*, 90(17):178103, 2003.

[38] ES Hoke, B Ross, and M Hoke. Auditory afterimage: tonotopic representation in the auditory cortex. *Neuroreport*, 9(13):3065–3068, 1998.

[39] AJ Norena and JJ Eggermont. Changes in spontaneous neural activity immediately after an acoustic trauma: implications for neural correlates of tinnitus. *Hearing Research*, 183(1):137–153, 2003.

[40] A Norena, C Micheyl, and S Chery-Croze. An auditory negative after-image as a human model of tinnitus. *Hearing Research*, 149(1):24–32, 2000.

[41] Q Summerfield, M Haggard, J Foster, and S Gray. Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics*, 35(3):203–213, 1984.

[42] JP Wilson. An auditory after-image. *Frequency Analysis and Periodicity Detection in Hearing*, 304, 1970.

[43] NF Viemeister and SP Bacon. Forward masking by enhanced components in harmonic complexes. *The Journal of the Acoustical Society of America*, 71(6):1502–1507, 1982.

[44] S Verhulst, T Dau, and CA Shera. Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission. *The Journal of the Acoustical Society of America*, 132(6):3842–3848, 2012.

[45] E Zwicker. Dependence of post-masking on masker duration and its relation to temporal effects in loudness. *The Journal of the Acoustical Society of America*, 75(1):219–223, 1984.

[46] MJ Penner. A power law transformation resulting in a class of short-term integrators that produce time-intensity trades for noise bursts. *The Journal of the Acoustical Society of America*, 63(1):195–202, 1978.

[47] DT Kemp. Otoacoustic emissions, travelling waves and cochlear mechanisms. *Hearing Research*, 22(1):95–104, 1986.

A Matlab code for creating the stimuli

```
clear all;
close all;

%sampling frequency
fs = 48000;
%highest frequency component in Hz
fh = 16000;
%length of the sound in seconds
soundlen = 60;

%% GENERATING ROSENBLITH PULSETRAIN
%Generate non-aliased Rosenblith type pulse train
%by summing cosine waves.

%pulse train fundamental frequency in Hz
f0 = 100;
%create time axis
t = 0:1/fs:soundlen;
%initialize sound
x = zeros(1,length(t));

%sum harmonic cosines up to fh to create the pulse train
for i=1:round(fh/f0)
    x = x + cos(2*pi*i*f0*t);
end

%normalize to the range [-1 1] and save to file
x = x./max(abs(x));

%save to vector
s1 = x;

%% GENERATING FREQUENCY MODULATED ROSENBLITH PULSETRAIN
% The same as above but adds frequency modulation by warping time

%modulation depth
mod_d = 0.01;
%modulation frequency
mod_freq = 4;
%modulation signal
mod = mod_d*sin(2*pi*mod_freq*t);
%warping the time with the modulation signal
t = t+mod;

%initialize sound
x = zeros(1,length(t));

%sum cosines to create the pulse train
for i=1:round(fh/f0)
    x = x + cos(2*pi*i*f0*t);
```

```
end

%normalize to the range [-1 1] and save to file
x = x./max(abs(x));
%save to vector
s2 = x;

%% GENERATING HARMONIC NOISE
% Generate harmonic noise (f0 and it's multiples) by summing ...
    components.

% The fundamental frequency
f0=100;

%create time axis
t = 0:1/fs:soundlen;
%initialize sound
x = zeros(1,length(t));

for i=1:round(fh/f0)
    x=x+cos(2*pi*f0*i*t+2*pi*rand);
end

%normalize
x = x./max(abs(x));

%save to vector
s3 = x;

%% GENERATING FILTERED NOISE
% Generate filtered white noise in the range 100-16000 Hz

%lower end of the range
f0=100;

%create time axis
t = 0:1/fs:soundlen;
%initialize sound
x=zeros(1,length(t));
%generate white noise
x=rand(size(x));

%high-pass and low-pass filter to only get the range 100-16000 Hz
[bp,ap]=butter(5,[f0/fs*2],'high');
x=filter(bp,ap,x);
[bp,ap]=butter(5,[16e3/fs*2],'low');
x=filter(bp,ap,x);

%normalize
x = x./max(abs(x));
%save to vector
s4 = x;

%% NORMALIZE ALL SOUNDS TO HAVE THE SAME RMS
```

```
norm_factor = min([rms(s1) rms(s2) rms(s3) rms(s4)]);

s1 = (norm_factor/rms(s1))*s1;
audiowrite('rosenblith_pulsetrain.wav',s1,fs);

s2 = (norm_factor/rms(s2))*s2;
audiowrite('rosenblith_pulsetrain_fm.wav',s2,fs);

s3 = (norm_factor/rms(s3))*s3;
audiowrite('harnoise.wav',s3,fs);

s4 = (norm_factor/rms(s4))*s4;
audiowrite('white_noise.wav',s4,fs);
```

B Adaptive tracks per condition

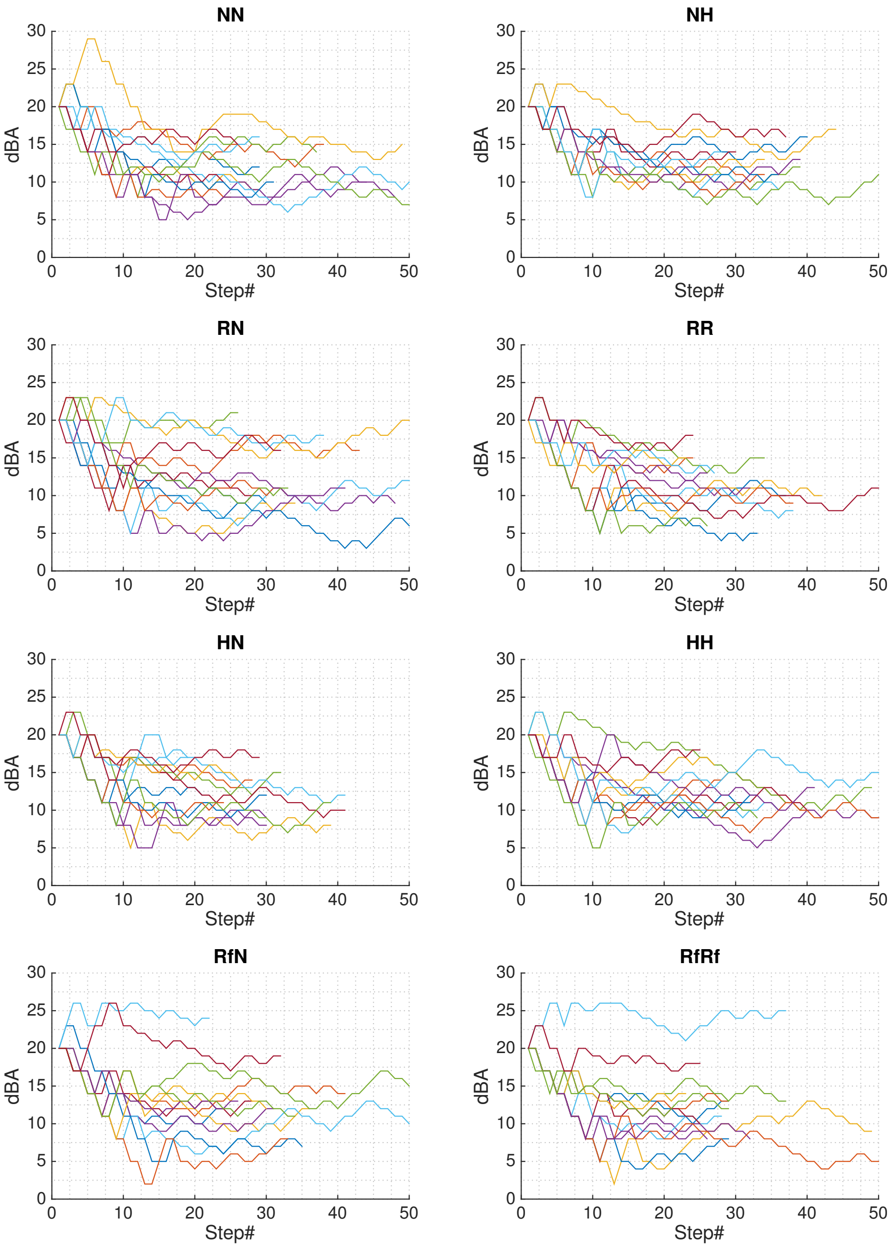

Figure B1: The adaptive tracks per condition.

C Adaptive tracks per subject

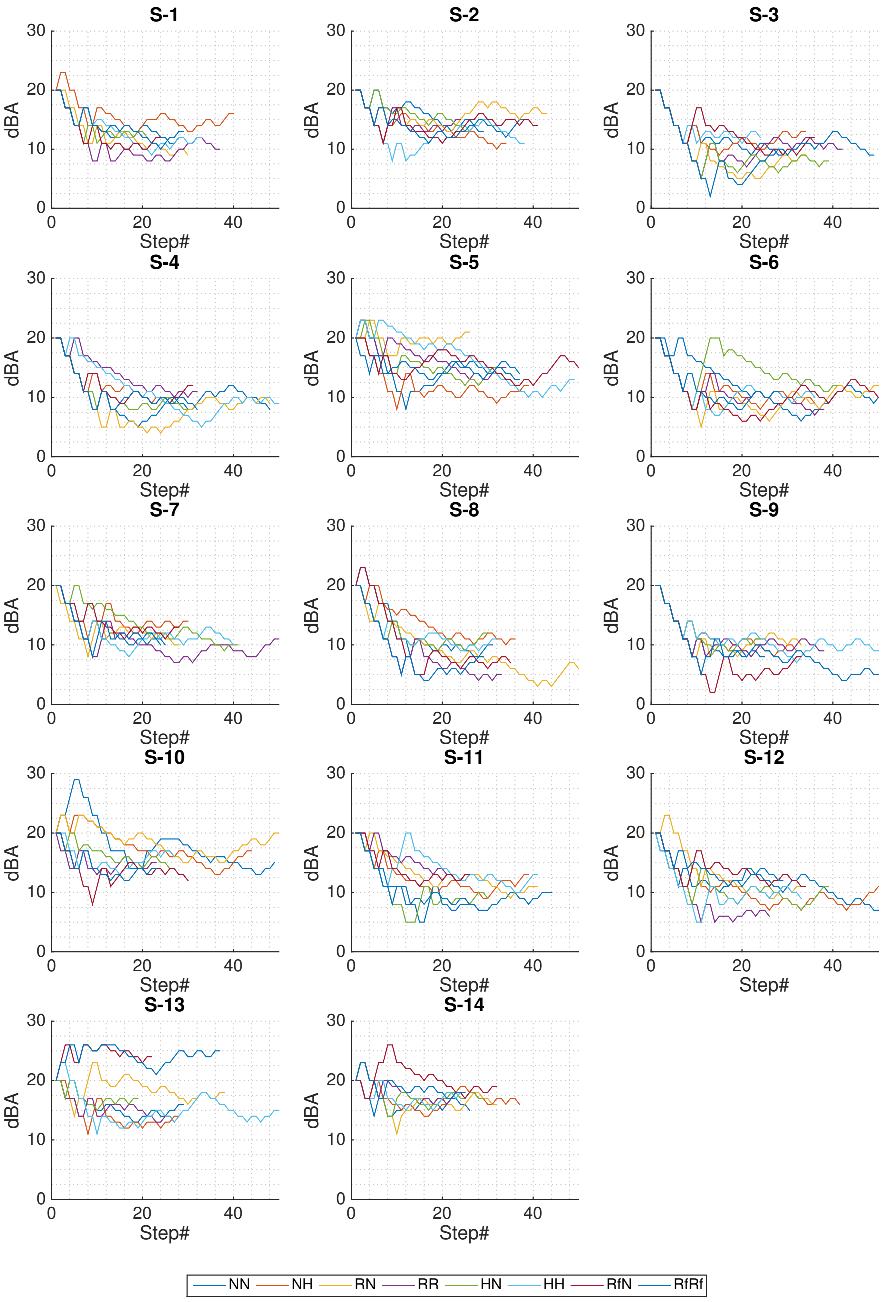

Figure C1: The adaptive tracks per subject.